

Hearing and saying

The functional neuro-anatomy of auditory word processing

C. J. Price,¹ R. J. S. Wise,¹ E. A. Warburton,¹ C. J. Moore,¹ D. Howard,² K. Patterson,³
R. S. J. Frackowiak¹ and K. J. Friston¹

¹Wellcome Department of Cognitive Neurology, Institute of Neurology, the ²Department of Psychology, Birkbeck College, London and the ³MRC Applied Psychology Unit, Cambridge University, UK

Correspondence to: Dr Cathy Price, Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1 3BG, UK.

Summary

The neural systems involved in hearing and repeating single words were investigated in a series of experiments using PET. Neuropsychological and psycholinguistic studies implicate the involvement of posterior and anterior left perisylvian regions (Wernicke's and Broca's areas). Although previous functional neuroimaging studies have consistently shown activation of Wernicke's area, there has been only variable implication of Broca's area. This study demonstrates that Broca's area is involved in both auditory word perception and repetition but activation is dependent on task (greater during repetition than hearing) and stimulus presentation (greater when hearing words at a slow rate). The peak of

frontal activation in response to hearing words is anterior to that associated with repeating words; the former is probably located in Brodmann's area 45, the latter in Brodmann's area 44 and the adjacent precentral sulcus. As Broca's area activation is more subtle and complex than that in Wernicke's area during these tasks, the likelihood of observing it is influenced by both the study design and the image analysis technique employed. As a secondary outcome from the study, the response of bilateral auditory association cortex to 'own voice' during repetition was shown to be the same as when listening to 'other voice' from a prerecorded tape.

Keywords: PET; listening; repeating; Broca's area; own voice

Abbreviations: BA = Brodmann's area; rCBF = regional cerebral blood flow

Introduction

The results from three new experiments and a re-analysis of two previously published experiments (Howard *et al.*, 1992; Price *et al.*, 1992) are used to investigate the perisylvian regions involved in the auditory perception and repetition of single words. The importance of the left perisylvian region for language is so well recognized that it has come to be referred to as the language organ (Pinker, 1994). The core regions are Wernicke's and Broca's areas (Mesulam, 1990). The boundaries of Wernicke's area are ill-defined but the minimum is the posterior part of the left superior temporal gyrus, while Broca's area is consistently defined as the posterior third of the left inferior frontal gyrus and considered to incorporate Brodmann's areas (BA) 44 and 45. Classically, damage to Wernicke's area is associated with poor word comprehension and relatively fluent speech, while damage to Broca's area results in relatively intact comprehension but poor speech fluency, at least at the single-word level. Lesions

to both are associated with poor repetition. The receptive/expressive, fluent/nonfluent dichotomies are useful clinically, but are not supported by more detailed analyses. Thus, speech perception is abnormal in patients with Broca's aphasia, and the articulation of speech sounds is subtly influenced by lesions of Wernicke's area (Caplan, 1987; Blumstein, 1994).

Several studies have been reported in which the brain areas involved in single word processing were investigated using PET. Activation in Wernicke's area was observed in response to hearing and repeating single words (Petersen *et al.*, 1988, 1989; Howard *et al.*, 1992). It was also observed when subjects made phonological judgements on heard syllables or non-words (Demonet *et al.*, 1992, 1994; Zatorre *et al.*, 1992). A region of the left posterior temporal lobe has also been found to respond in a qualitatively different way to the rest of the auditory cortex when subjects listen to words (Price *et al.*, 1992) and extrasylvian regions in the left

ventral temporal and posterior, inferior parietal lobes are engaged when subjects retrieve words from semantic memory (Martin *et al.*, 1995; Warburton *et al.*, 1996).

Activation of the left inferior frontal region (Broca's area) during auditory word processing has been observed during phonological judgements on heard stimuli (Demonet *et al.*, 1992, 1994; Zatorre *et al.*, 1992), the retrieval of words (Petersen *et al.*, 1988, 1989; Wise *et al.*, 1991; Warburton *et al.*, 1996), semantic judgements on heard words (Demonet *et al.*, 1994; Warburton *et al.*, 1996) and silent 'repetition' of non-words with three repetitions per stimulus (Warburton *et al.*, 1996). However, these tasks all require auditory-verbal short-term memory to remember the auditory stimuli whilst phonological or semantic decisions are made, related words are retrieved or whilst non-words are silently 'repeated' three times. Since activity in Broca's area increases during auditory-verbal short-term memory tasks (Paulesu *et al.*, 1993), its role during these tasks may be attributable to phonological rehearsal.

A role for Broca's area during perception and articulation of heard words that can be distinguished from parallel processes such as auditory-verbal short-term memory has been shown less consistently. In a posterior region of Broca's area (BA 44), increased activation during repetition was detected when the contrast was with listening to words at the same rate of presentation (Petersen *et al.*, 1988, 1989) but not in contrast to saying 'crime' in response to hearing nonphononetic acoustic stimulation (Howard *et al.*, 1992). In the anterior region (near the border of BA 45 and the anterior insula) increased activity was detected during auditory detection tasks on stimuli with rapid temporal changes (e.g. words, syllables and tone sequences) when the baseline was silent visual fixation (Fiez *et al.*, 1995) but not during auditory detection on steady state vowels or during passive listening (Petersen *et al.*, 1988, 1989; Fiez *et al.*, 1995). Fiez *et al.* (1995) interpret these findings in terms of a role for the frontal operculum in rapid temporal processing which underlies some forms of phonological analysis and is not required for listening to and identifying the sensory input. However, activation of BA 45 is detected during passive listening when the rate of heard word presentation is slow (Warburton *et al.*, 1996).

As lesion data strongly supports the notion that the left inferior frontal gyrus is involved in auditory word perception and production, it seemed worth pursuing changes in experimental design and the use of more sensitive PET scanning techniques and image analysis to investigate whether Broca's area as well as Wernicke's area could reliably be associated with hearing and repeating word lists. Although experimental design is paramount, methodological factors play an important role. Recent advances in PET instrumentation have increased the sensitivity and resolution of the scanners; e.g. three-dimensional PET offers approximately a six-fold increase in coincidence count rates over two-dimensional methods (which outweighs the disadvantages of increased scattered and random event rates) (Bailey *et al.*,

1993). The recent developments in data analysis are addressed in the Methods section.

The new image processing software (Friston *et al.*, 1994, 1996) was used to analyse results from three new studies and those from two previously published studies (Howard *et al.*, 1992; Price *et al.*, 1992) to investigate further hearing and repeating single words. The experimental designs are summarized in the Methods section. The purpose was to identify those regions involved in auditory word perception and repetition, with particular emphasis on posterior and anterior left perisylvian cortex. As a number of tasks involved repetition and one involved reading words aloud, there was the opportunity to study articulation and the response in the auditory cortex to the sound of 'own' voice during speaking. The activation tasks consisted of hearing, repeating or reading word lists at various rates of presentation. The control states, with which the activation tasks were contrasted, varied from rest (eyes closed, no stimuli, no response) to listening to words played backwards (to subtract the signal from the early acoustic processing of words from that due to later featural and lexical processing) and saying the same word in response to each stimulus (to subtract signals due to articulation).

Methods

Subjects

In all experiments, the subjects were English speaking, right-handed volunteers with no history of neurological disorders. The study was approved by the local ethical committee and each subject gave informed consent to undergo the PET scans. The experimental designs are summarized below to specify the activation and baseline conditions, the number and ages of the subjects, whether the data acquisition was two-dimensional or three-dimensional (*see* Data acquisition), the order of conditions across the 6–12 scans, the contrasts used in data analysis and the regions identified by the contrast.

Experimental design

Experiment 1 (reanalysis of the data from Howard *et al.*, 1992)

Activation condition	(A) Repeating aloud heard words at 40 words per minute (w.p.m.)
Baseline condition	(B) Listening to reversed words at 40 w.p.m. and saying 'crime'
Subjects	Twelve (aged 18–70 years)
Acquisition	Two-dimensional, three scans per subject, 24 activation scans, 12 baseline scans
Condition order	Latin square design
Contrast	A-B
Regions identified	(i) Auditory word perception ('own' plus 'other' voice) with

subtraction of early acoustic processing
(ii) Speech production during repetition with subtraction of articulation

Experiment 2

Activation condition (A) Repeating aloud heard words at 40 w.p.m.
 Baseline condition (B) Listening to reversed words at 40 w.p.m. and saying 'crime'
 Subjects Four (aged 28–62 years)
 Acquisition Three-dimensional, 12 scans per subject, 24 scans per condition
 Condition order ABABABABABAB
 Contrast A-B
 Regions identified (i) Auditory word perception ('own' plus 'other' voice) with subtraction of early acoustic processing
 (ii) Speech production during repetition with subtraction of articulation

Experiment 3

Activation condition (A) Listening to words at 40 w.p.m.
 Baseline condition (B) Listening to reversed words at 40 w.p.m.
 Subjects Four (aged 35–62 years)
 Acquisition Three-dimensional, 12 scans per subject, 24 scans per condition
 Condition order ABABABABABAB
 Contrast A-B
 Regions identified (i) Auditory word perception ('other' voice) with subtraction of early acoustic processing

Experiment 4

Activation condition (A1) Listening to words at 40 w.p.m.
 Activation condition (A2) Repeating words at 40 w.p.m.
 Activation condition (A3) Repeating words at 20 w.p.m.
 Activation condition (A4) Reading aloud words at 40 w.p.m.
 Activation condition (A5) Reading silently words at 40 w.p.m.
 Baseline condition (B) Resting, eyes closed, silence.
 Subjects Six (age 22–33 years)
 Acquisition Two-dimensional, six scans per subject, six scans per condition

Condition order Latin square design
 Contrasts (1) A1–B
 (2) A2–B and A3–B
 (3) A4–B and A5–B
 (4) A2–A1 and A3–A1 and
 (5) A4–A1
 Regions identified (1) Auditory word perception ('other' voice only)
 (2) Auditory word perception ('own' plus 'other' voice)
 (3) Visual word perception and speech production during reading
 (4) Speech production during repetition
 (5) Speech production during reading

Experiment 5 (reanalysis of the data from Price et al., 1992)

Activation condition (A1) Listening to words at 10 w.p.m.
 Activation condition (A2) Listening to words at 30 w.p.m.
 Activation condition (A3) Listening to words at 50 w.p.m.
 Activation condition (A4) Listening to words at 70 w.p.m.
 Activation condition (A5) Listening to words at 90 w.p.m.
 Baseline condition (B) Resting, eyes closed, silence
 Subjects Six (age 24–49 years)
 Acquisition Two-dimensional, six scans per subject, six scans per condition
 Condition order Latin square design
 Contrasts (1) A2 + A3 – B
 (2) [10 × A1] + [30 × A2] + [50 × A3] + [70 × A4] + [90 × A5]
 (3) [2 × A1] – A5 – B
 Regions identified (1) Auditory word perception ('other' voice)
 (2) Linear increases with increased rate of hearing words
 (3) Non-linear increases with increased rate of hearing words where the greatest increase is between rest and 10 w.p.m., i.e. (A1–B) > (A5–A1)

Data acquisition

In the experiments with two-dimensional acquisition, subjects were studied using a Siemens 931–08/12 PET scanner (Spinks et al., 1988). Regional cerebral blood flow (rCBF) was measured using a C¹⁵O₂ inhalation technique (Lammertsma

et al., 1990). During each 3.5 min scan, the subject inhaled $C^{15}O_2$ at a concentration of 6 MBq ml^{-1} and a flow rate of 500 ml min^{-1} through a standard oxygen face mask for a period of 2 min. Intervals between scans lasted 12–15 min.

In the experiments with three-dimensional acquisition, subjects were studied using a CTI-Siemens 953B dedicated head scanner (Spinks et al., 1992). A venous cannula to administer the tracer was inserted in a right antecubital fossa vein. Approximately 550 MBq of $H_2^{15}O$ in 3 ml of normal saline was loaded into the intravenous tubing over 20 s and flushed into the subject over 20 s at a rate of 10 ml min^{-1} by an automatic pump. After a delay of $\sim 35 \text{ s}$, a rise in counts could be detected in the head which peaked 30–40 s later (depending on individual circulation time). The interval between successive $H_2^{15}O$ administrations was 10 min. The data were acquired in one 90 s frame, beginning 0–5 s before the rising phase of the head curve (Silbersweig et al., 1993).

For both experiments, correction for attenuation was made by performing a transmission scan with an exposed $^{68}\text{Ge}/^{68}\text{Ga}$ external source at the beginning of each study. Images were reconstructed by filtered back projection (Hanning filter, cut-off frequency 0.5 cycles per pixel), giving a transaxial resolution of 8.5 mm full width at half maximum and displayed in a 128×128 pixel format with 43 planes creating approximately cubic voxels. In the axial dimension the resolution was 8 mm for the Siemens 931–08/12 scanner and 6 mm for the CTI-Siemens 953B scanner. However, the images from both scanners were subsequently smoothed to a resolution of 16 mm (see below).

Image and statistical analysis

Image analysis was performed on a SPARC STATION II (Sun Microsystems Europe Inc., Surrey, UK), using an interactive image display software package (ANALYZE, Biodynamic Research Unit, Mayo Clinic, Rochester, Minn., USA; Robb and Hanson, 1991) and statistical parametric mapping (Friston et al., 1991b, 1996; SPM software, Wellcome Department of Cognitive Neurology, London, UK). Calculations and image matrix manipulations were performed in PRO MATLAB (Mathworks Inc., Sherborn, Mass., USA).

The relative distribution of rCBF in the brain was indexed by the accumulated counts over the scanning period, which reliably reflects flow in the physiological range (Fox and Mintun, 1989; Mazziotta et al., 1985). These data were reorientated parallel to the intercommissural line, then standardized for brain size and shape (Friston et al., 1996). When stereotactically normalized, one voxel in the transformed image represented 2 mm in the x and y dimensions and 4 mm in the z dimension, which corresponds to the dimensions of the atlas of Talarach and Tournoux (1988). These images were smoothed using an isotropic 16 mm kernel to account for variation in gyral anatomy and individual variability in structure–function relationships, and to improve the signal-to-noise ratio.

Global variance between conditions was removed using

analysis of covariance (ANCOVA). This analysis generated maps of adjusted mean values of rCBF for each of the experimental conditions with corresponding associated adjusted error variance maps required for comparison of these means. The differences between conditions were assessed by formal comparisons of the condition specific rCBF maps. Adjusted condition means and variances were compared on a voxel-by-voxel basis by weighting the condition means by an appropriate contrast (Friston et al., 1991b). The resulting map of Z values constituted the statistical parametric map.

Differences in data analysis from that used in previous publications

Realignment

Originally realignment permitted a scan-to-scan readjustment of pitch and z translation (two degrees of freedom) based on a least squares match to a series of coronal template or ‘model’ slices. The advent of more comprehensive realignment techniques using six degrees of freedom (Woods et al., 1992) indicated the importance of correcting more completely for subject movements between scans. We now use a computationally efficient least squares technique with rigid-body affine transformations to minimize the sum of squares between each scan and the first. This approach can be shown to have a precision of 10–100 μm for some functional MRI data (Friston et al., 1996).

Spatial normalization

The original spatial normalization procedure (Friston et al., 1991a) used a series of linearizing devices and a crude nonlinear ‘warping’ in one dimension. This nonlinear transformation was applied to each radius of every transverse slice, in order to match the observed image to a standard template. Current techniques are simpler and more comprehensive, using a simultaneous solution for a 12 parameter affine and a six parameter nonlinear (second order) spatial transformation in three dimensions. This solution is obtained using standard least squares after linearizing the problem with a first order Taylor expansion. Fine scale nonlinear differences in local anatomy that remain are removed using a set of smooth Fourier-like basis functions to transform each transverse section. This approach is non-interactive and therefore reliable (Friston et al., 1996).

Smoothing

The improvement in co-localization of homologous anatomical structures (within and between subjects) that ensues from the above techniques means less smoothing of the data is needed. Our original analyses were based on an anisotropic smoothing with a $20 \times 20 \times 12 \text{ mm}$ full width half maximum Gaussian kernel. This has been optimized as an isotropic 16 mm kernel on two-dimensional data and less for three-

Table 1 Results from repeating single words: contrasted with rest, and saying a prespecified word in response to hearing reversed words

		Left	Medial	Right
<i>Repeating versus rest (Experiment 4)</i>				
Post. superior temporal (BA 21/22)	20 w.p.m.	-54, -44, +12: 4.6		+46, -38, +8: 5.8
	40 w.p.m.	-52, -36, +12: 5.3		+50, -40, -12: 6.9
Mid superior temporal (BA 41/42/22)	20 w.p.m.	-56, -16, +8: 4.8		+56, -22, 0: 4.7
	40 w.p.m.	-56, -16, +8: 6.4		+56, -22, 0: 5.3
Mid middle temporal (BA 21/22)	20 w.p.m.	-60, -4, +8: 3.7		+56, -8, +4: 3.4
	40 w.p.m.	-56, -2, 0: 4.6		+56, -8, +4: 4.9
Central sulcus (BA 3/4)	20 w.p.m.	-48, -10, +24: 3.2		NS
	40 w.p.m.	-46, -12, +32: 4.8		+28, -14, +32: 3.7
Insula	20 w.p.m.	NS		+26, -2, +16: 3.7
	40 w.p.m.	-32, +4, +4: 3.1		+42, -14, +8: 3.6
	40 w.p.m.			+28, +14, +8: 3.6
Cerebellar vermis	20 w.p.m.		+4, -58, -8: 3.1	
	40 w.p.m.		-2, -52, -8: 3.5	
Thalamus/lentiform nucleus	20 w.p.m.	-22, -22, +8: 3.7	-14, -4, +12: 3.1	
	20 w.p.m.		-8, -24, 0: 3.3	
	40 w.p.m.	-22, -18, +8: 4.7	-6, -18, +4: 4.6	
Supplementary motor area (BA 6)	20 w.p.m.		NS	
	40 w.p.m.		-6, -4, +60: 3.3	
<i>Repeating (40 w.p.m.) versus saying 'crime' to reversed words (Experiments 1 and 2)</i>				
Posterior middle temporal (BA 21)	Expt 2	-52, -52, +4: 4.5		
Inferior temporal (BA 20/21)	Expt 1	-56, -44, -8: 3.4		
	Expt 1	-46, -34, -12: 3.7		
	Expt 2	-48, -24, -8: 4.2		
Inferior frontal gyrus (BA 44)	Expt 1	-42, +12, +12: 3.1		
	Expt 2	-48, +14, +12: 4.7		
Inferior frontal sulcus (BA 9/44)	Expt 1	-40, +12, +28: 3.2		
Central sulcus/precentral sulcus	Expt 1	-42, -2, +40: 3.1		
	Expt 2	-48, -20, +36: 4.3		
Cerebellar vermis	Expt 1		4, -82, -28: 3.6	
	Expt 2		+6, -58, -28: 3.0	
Lingual gyrus	Expt 1		-2, -64, 0: 3.5	
	Expt 1		-4, -96, -4: 3.3	
Thalamus	Expt 2			+20, -18, +12: 4.0
Medial parietal (paracentral lobe)	Expt 2		-4, -28, +56: 3.2	
Anterior cingulate gyrus (BA 32)	Expt 1		0, +18, +28: 4.1	
	Expt 2		-6, +22, +36: 3.7	
Supplementary motor area (BA 6)	Expt 2		-6, 0, +56: 6.3	

For explanation of data in this table *see* footnotes to Table 4.

dimensional data and single subject experiments. The reduction in smoothing gives better anatomical resolution in statistical images or parametric maps.

Statistical inference

Substantial advances have been made in the application of the theory of Gaussian fields to statistical parametric maps (Friston *et al.*, 1991b, 1994; Worsley *et al.*, 1992). A *P* value can be assigned to an activation focus that has been corrected for the entire volume of the statistical parametric map analysed. However, the hypotheses investigated in this paper have been framed in terms of specified anatomical locations, removing the need to correct for multiple comparisons; only the *Z* values and uncorrected *P* values are reported.

Results

The results of the five different experiments identify activations associated with (i) repeating words (Experiments 1, 2 and 4; Table 1, Fig. 1A), (ii) listening to words (Experiments 3, 4 and 5; Table 2, Fig. 1B), (iii) repeating versus listening (Experiment 4; Table 3, Fig. 1C), and (iv) reading aloud versus listening (Experiment 4; Table 4, Fig. 1D).

Perisylvian regions

Mid-dorsal temporal activations

Listening to words contrasted with rest produced bilateral temporal lobe activations (with the peaks in association

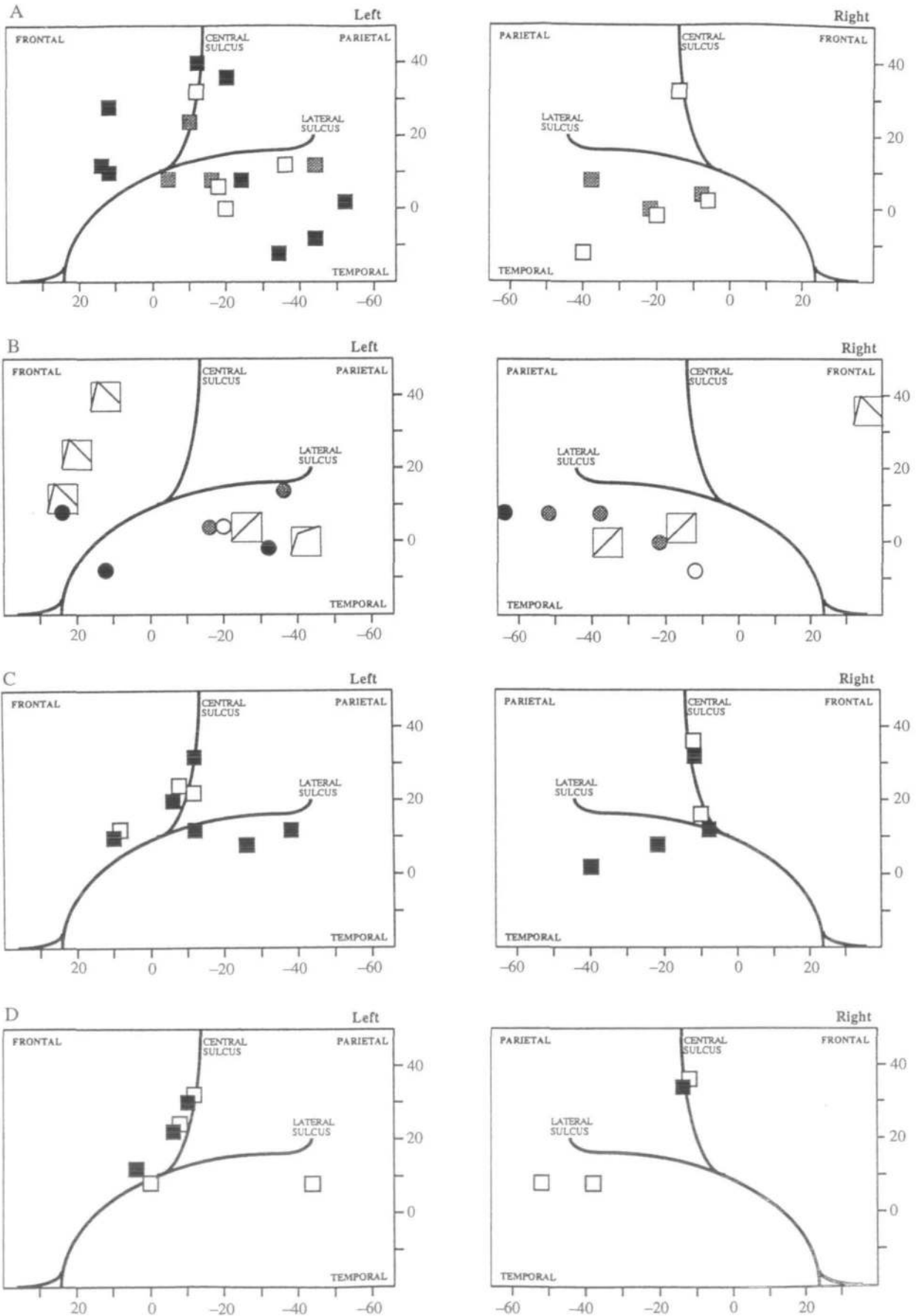


Table 2 Results from listening to single words spoken: by the voice of another contrasted with rest, hearing reversed words, and reading aloud

	Left	Medial	Right
<i>Listening versus rest (Experiments 4 and 5)</i>			
Mid superior temporal (BA 22)	Expt 4 -60, -16, +4: 5.1		+56, -22, 0: 4.2
	Expt 5 -54, -26, +4: 6.0		+62, -18, +4: 5.2
Post. superior temporal (BA 21/22)	Expt 4 -52, -36, +12: 3.9		+50, -52, +8: 3.2
	Expt 4		+48, -38, +8: 5.2
	Expt 5 -54, -42, 0: 6.3		+60, -48, +4: 3.1
	Expt 5		+52, -34, -4: 4.8
<i>Linear relationships between rCBF and presentation rate of words (Experiment 5)</i>			
Mid superior temporal (BA 21/22)	-54, -26, +4: 8.1		+56, -16, +4: 7.5
Post. middle temporal (BA 21)			+56, -36, 0: 6.8
<i>Non-linear relationships between rCBF and presentation rate of words (Experiment 5)</i>			
Post. middle temporal (BA 21)	-52, -42, 0: 6.0		
Middle frontal gyrus (BA 9/46)	-44, +12, +40: 3.4		+44, +36, +24: 3.8
Inferior frontal gyrus (BA 45)	-50, +20, +24: 3.7		
	-46, +24, +12: 3.2		
<i>Listening 40 w.p.m. versus reversed words 40 w.p.m. (Experiment 3)</i>			
Post. inferior temporal sulcus (BA 37/39)			+50, -64, +8: 3.3
Post. middle temporal gyrus (BA 21)	-54, -32, -4: 4.1		
Ant. superior temporal gyrus (BA 38)	-46, +12, -8: 3.5		
Inferior frontal gyrus (BA 45)	-48, +24, +8: 3.7		
<i>Listening (40 w.p.m.) versus reading aloud 40 w.p.m. (Experiment 4)</i>			
Mid superior temporal gyrus (BA 21/22)	-52, -20, +4: 3.4		+48, -12, -8: 3.4

For explanation of data in this table see footnotes to Table 4.

cortex). Activity was linearly related to the rate of word presentation (Price *et al.*, 1992; Table 2, Experiment 5) but was not detected when acoustic stimulation was removed in the contrast with hearing reversed words (Table 2, Experiment 3).

Similarly, when subjects were repeating words, there were bilateral temporal lobe activations if the contrast was with rest but not if the contrast was with hearing reversed words (Table 1). The auditory input during repeating is a mixture of own and other voice; during repeating 20 w.p.m., the total auditory input is 40 w.p.m. (20 w.p.m. other voice plus 20 w.p.m. own voice), during repeating 40 w.p.m. the total auditory input is 80 w.p.m. (40 w.p.m. other voice plus 40 w.p.m. own voice). When the repeating conditions were contrasted with listening to 40 w.p.m., bilateral temporal lobe activations were revealed for repeating 40 w.p.m. but not for

repeating 20 w.p.m. demonstrating that hearing 'own' voice is additive with hearing 'other' voice, at least during repeating. It was not quite the case during reading aloud, as the contrast of listening to 40 w.p.m. with reading aloud 40 w.p.m. did show small bilateral mid-temporal activations (Table 2).

Posterior ventral temporal activations

In the posterior and inferior region of the left temporal lobe, activations during listening and repeating were detected in contrast to both rest and reversed words illustrating specificity to word processing in this area. For listening there was also a small right inferior lateral temporal activation in contrast to reversed words that was not detected for repeating. Reanalysis of the study by Price *et al.* (1992) confirmed the nonlinear response in the left posterior temporal region where

Fig. 1 Each two-part section of this figure is a simplified representation of the left and right perisylvian region, including the ventral half of the central sulcus and the whole extent of the lateral sulcus. The y and x coordinates (in mm) from the atlas of Talarach and Tournoux (1988) are on the horizontal and vertical axes, respectively (the x coordinate is not depicted on these two dimensional representations)—0 is on the plane of the anterior commissure. Lateral cortical activations have been plotted on these representations to permit the perisylvian regions involved in listening and repeating to be more readily appreciated than from Table 1. (A) Repeating: hatched squares = repeating 20 w.p.m. versus rest; open squares = repeating 40 w.p.m. versus rest; filled squares = repeating 40 w.p.m. versus saying the same prespecified word in response to hearing reversed words. (B) Listening: hatched circles = listening 40 w.p.m. versus rest; closed circles = listening 40 w.p.m. versus listening reversed words 40 w.p.m.; open circles = listening 40 w.p.m. versus reading aloud 40 w.p.m. Straight line graph in square = regions where rCBF increased linearly with the rate of hearing single words; other graphs in squares = regions where the relationship between rCBF and the rate of hearing single words was nonlinear. (C) Repeating: open squares = repeating 20 w.p.m. versus listening 20 w.p.m.; closed squares = repeating 40 w.p.m. versus listening 40 w.p.m. (D) Reading: open squares = reading aloud 40 w.p.m. versus rest; closed squares = reading aloud 40 w.p.m. versus listening 40 w.p.m.

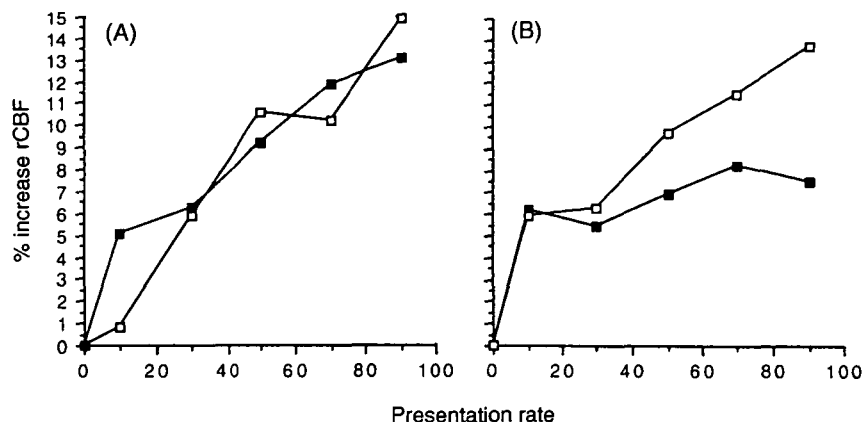


Fig. 2 Graphs of percentage increase in rCBF relative to rest as word presentation increases from rest (0 w.p.m.) to 90 w.p.m. (Experiment 4) in the left and right mid- (A) and posterior (B) temporal regions. Filled squares = left hemisphere; open squares = right hemisphere.

the main increase was between 0 and 10 w.p.m., with little further increase between 10 and 90 w.p.m. (Table 2). Figure 2 illustrates that the percentage rCBF increase in the left posterior temporal lobe plateaued as the rate of heard words increased, at a level below that observed in the right homologous region, where activity continued to increase with increasing rate of hearing words. Nevertheless, when repetition was contrasted with listening to 40 w.p.m. there was an additional increase in activation in the left posterior temporal lobe (Experiment 4) which suggests that there is an effect of speech production that is additive to the effect of speech perception.

Left inferior frontal gyrus

Posterior Broca's area activation (BA 44) was seen during repeating if the control state was listening to words or saying 'crime' in response to reversed words but not when the control state was rest (Tables 1 and 3; Fig. 1A–C). Activation close to posterior Broca's area (in BA 6, the precentral sulcus) was also associated with reading aloud at 40 w.p.m. (see Table 4).

Anterior Broca's area activation (BA 45) was detected during listening to words when the contrast was with hearing reversed words and at slow rates of presentation when the contrast was with rest (Table 2, Figs 1B and 3). This nonlinear response to increasing rate of heard words was also observed in bilateral middle frontal gyri (Fig. 3).

Bilateral anterior insular activations were seen in the contrasts of repeating with rest (Table 1) and listening to words (Table 3) and in the left anterior insula during reading aloud contrasted with listening to words (Table 4).

Sensorimotor cortex, supplementary motor area, subcortical grey matter and midline cerebellum

These regions were only activated when articulation occurred (Tables 1, 2 and 4). They were also more active when

different words were repeated compared with the more 'automatic' speech task of saying the same word repeatedly in response to auditory stimuli (Table 1). In most of the contrasts there were bilateral sensorimotor activations; in those contrasts when it was unilateral, it was on the left. The lentiform nucleus and thalamic activations were mostly on the left, the cerebellar activations were in the midline.

Differences between original findings and reanalysis of data from Howard et al. (1992) and Price et al. (1992)

In the original analysis of these studies no frontal lobe activations were detected. The more sensitive analysis reported in this paper reveals that Broca's area is active during repetition [reanalysis of Howard *et al.* (1992) and Experiment 2] and in a complex fashion during listening [reanalysis of Price *et al.* (1992) and Experiment 3].

Discussion

Areas involved in speech perception

Perception of another voice

Speech perception involves acoustic, phonological and lexical semantic processing of heard words. It was assumed in this study that listening to real words activates all these processes, whereas listening to reversed words only activates acoustic processes (as phonological features were largely absent from the stimuli). Hearing and repeating words in contrast to rest activated bilateral dorsal temporal regions with the peaks of activation located in association cortex. These regions are associated with acoustic processing because activation is largely abolished when the contrast is with listening to reversed words. Areas that were more active for words than reversed words were located in the left posterior temporal lobe with peaks of activation in middle temporal gyrus, inferior temporal sulcus and even in the inferior temporal

Table 3 Results from repeating (20 w.p.m. or 40 w.p.m.): contrasted with listening (40 w.p.m.) (Experiment 4)

	Left	Medial	Right
Post. sup/middle temporal (BA 21/22)	20 w.p.m.	-62, -44, +8: 2.6	NS
	40 w.p.m.	-52, -38, +12: 3.0	+50, -40, +12: 5.1
	40 w.p.m.	-60, -26, +8: 3.7	
Mid superior temporal (BA 22/42)	20 w.p.m.	NS	NS
	40 w.p.m.		+60, -22, +8: 3.2
	40 w.p.m.	-60, -12, +12: 4.6	
Temporal operculum	20 w.p.m.	NS	NS
	40 w.p.m.	NS	+46, +2, -8: 3.5
	40 w.p.m.	-54, -8, +24: 4.2	+48, -12, +36: 4.6
Central sulcus (BA 4)	20 w.p.m.	-46, -12, +32: 5.2	+46, -12, +32: 5.4
	40 w.p.m.	-44, -12, +24: 3.9	+58, -10, +16: 3.8
	40 w.p.m.	-56, -6, +20: 4.9	+56, -8, +12: 4.4
Insula/lentiform nucleus	20 w.p.m.	NS	+28, 0, +16: 3.3
	40 w.p.m.	-30, -12, +16: 4.1	+36, -6, 0: 3.6
	40 w.p.m.	-26, -4, +4: 4.5	+26, -6, +16: 5.1
	40 w.p.m.	-32, +4, +4: 4.2	
	40 w.p.m.	-58, +8, +12: 3.9	
Inferior frontal gyrus (BA 44)	20 w.p.m.	-58, +8, +12: 5.2	
	40 w.p.m.		
Cerebellar vermis	20 w.p.m.		+6, -58, -12: 4.0
	40 w.p.m.		+4, -58, -12: 4.0
Thalamus/lentiform nucleus	20 w.p.m.	-24, -10, +4: 3.9	
	40 w.p.m.	-18, -16, +8: 4.0	
Supplementary motor area (BA 6)	20 w.p.m.		-4, -4, +60: 3.0
	40 w.p.m.		-6, -4, +60: 4.3

For explanation of data in this table see footnotes to Table 4.

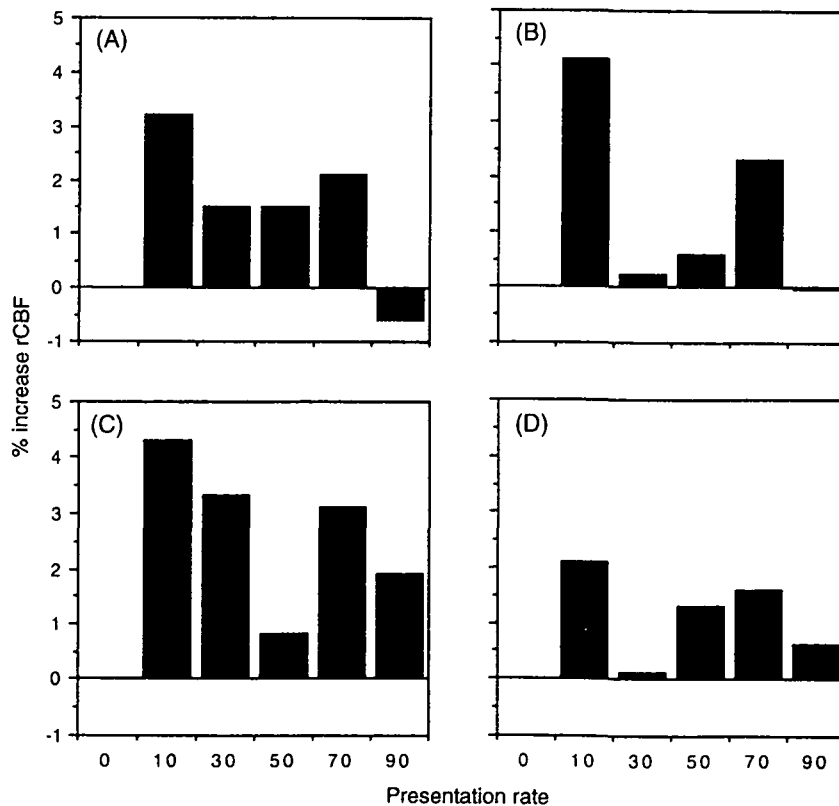


Fig. 3 Histograms of percentage increase in rCBF relative to rest as word presentation increases (Experiment 4) in left and right prefrontal regions: A, left inferior frontal (at $z = +12$); B, left inferior frontal ($z = +24$); C, right middle frontal; D, left middle frontal.

Table 4 Results from reading silently and aloud: contrasted with rest, and reading aloud contrasted with listening to the voice of another

	Left	Medial	Right
<i>Reading silently 40 w.p.m. versus rest (Experiment 4)</i>			
Post. inferior temporal sulcus (BA 37)			+48, -58, +4: 3.1
Posterior superior temporal sulcus (BA 21/22)	-60, -46, +12: 2.9		
Parieto-occipital junction (BA 19/39)	-32, -54, +32: 3.3		+40, -54, +32: 3.0
<i>Reading aloud 40 w.p.m. versus rest (Experiment 5)</i>			
Post. middle/superior temporal (BA 21/22)	-58, -44, +8: 4.1		+50, -52, +8: 4.0
Central sulcus	-42, -12, +32: 5.5		+46, -38, +8: 5.3
Central sulcus (sylvian)	-54, -6, +24: 4.7		+44, -12, +36: 4.4
Cerebellar vermis	-58, 0, +8: 4.4	+6, -62, -8: 4.3	
Thalamus/lentiform nucleus	-22, -22, +8: 4.1	+16, -26, +12: 4.0	
Anterior thalami		-2, -16, +4: 4.7	
Supplementary motor area (BA 6)		-2, -2, +56: 3.1	
<i>Reading aloud 40 w.p.m. versus listening 40 w.p.m. (Experiment 4)</i>			
Central sulcus (BA 4)	-46, -12, +32: 5.8		+48, -12, +36: 5.8
	-56, -6, +24: 5.4		
Precentral gyrus (BA 6)	-56, +4, +12: 5.5		
Insula	-36, +6, +4: 3.3		
Cerebellar vermis		+6, -62, -12: 4.9	
Thalami/lentiform nucleus	-26, -12, +8: 3.8	0, -10, +4: 3.8	
Supplementary motor area) (BA 6)		-4, 0, +60: 3.9	

All the comparisons were categorical (A versus B) except those in Experiment 4, which was designed to correlate rCBF increases with rates of word presentation. The anatomical region where each activation was observed is associated with its Brodmann's area (BA), the *x*, *y* and *z* coordinates (in millimetres) of the peak *z* score within each activated region and the *z* score in bold, e.g. (BA 21) (-54, -42, 0: **6.3**). The Brodmann's areas are only approximate as the boundaries of architectonic zones are individually variable in relation to sulci and gyri. Temporal regions are subdivided into posterior (post.), mid and anterior (ant.): posterior is defined as behind the coronal plane of the primary auditory cortex (*y* < -24 mm); mid as between the primary auditory cortex and the coronal plane of the anterior commissure (*y* < 0 > -24 mm); and anterior as the temporal lobe in front of the anterior commissure (*y* > 0 mm). For each contrast the activated regions are grouped into left and right lateral cortical areas (including opercular and insular regions); activations in medial cortical regions, basal ganglia, thalamic nuclei and the cerebellum are included together as a medial group. Activation in visual areas is not reported for the reading contrasts.

gyrus. This is in accordance with the notion that the the left posterior perisylvian and extrasylvian cortices are associated with lexical and semantic processing of heard words.

The left inferior frontal gyrus is also activated by hearing words. Precise localization is not possible using anatomical normalization of group data into the stereotactic space of the atlas of Talarach and Tournoux (1988), but the activation peak appears to lie in the anterior part of Broca's area (BA 45). The response is detected in contrast to attending to reversed words but is not always observed when rest is used as the control state possibly because the so-called rest state is associated, to a variable degree between subjects, with activities that are likely to involve frontal areas such as self-initiated thoughts and inner speech. Across a group of subjects this might result in variance in regional perfusion that would mask a consistent but subtle increase in activity in the left inferior frontal gyrus associated with hearing words. Attention to another auditory input (reversed words) allowed the left frontal response to hearing words to be observed. Another factor was the rate at which words were presented, the

reanalysis of the study of Price *et al.* (1992) showed a complex response in anterior Broca's area (and other left and right frontal regions) to words presented at varying rates—the greatest response was to the slowest rate of presentation of words, consistent with the findings by Warburton *et al.* (1996) that significant activation in anterior Broca's area is found when hearing words at a rate as slow as 4 w.p.m. is contrasted with rest. As the frontal regions shown in Fig. 3 are not demonstrating a response understandable in terms of the integration of phasic responses to the external stimuli (heard words) over the period of acquisition of data, it is not possible to understand what attribute of the input these regions are processing. The magnitude of the tonic increase in activity in these regions seems to be inversely related (not necessarily linearly) to the tempo of the external stimuli. Further studies will be needed to clarify whether this response is restricted to linguistically meaningful stimuli or whether it is also differentially sensitive to a varying tempo of simpler acoustic stimuli. Figure 4 illustrates the neural correlates of word perception and word production.

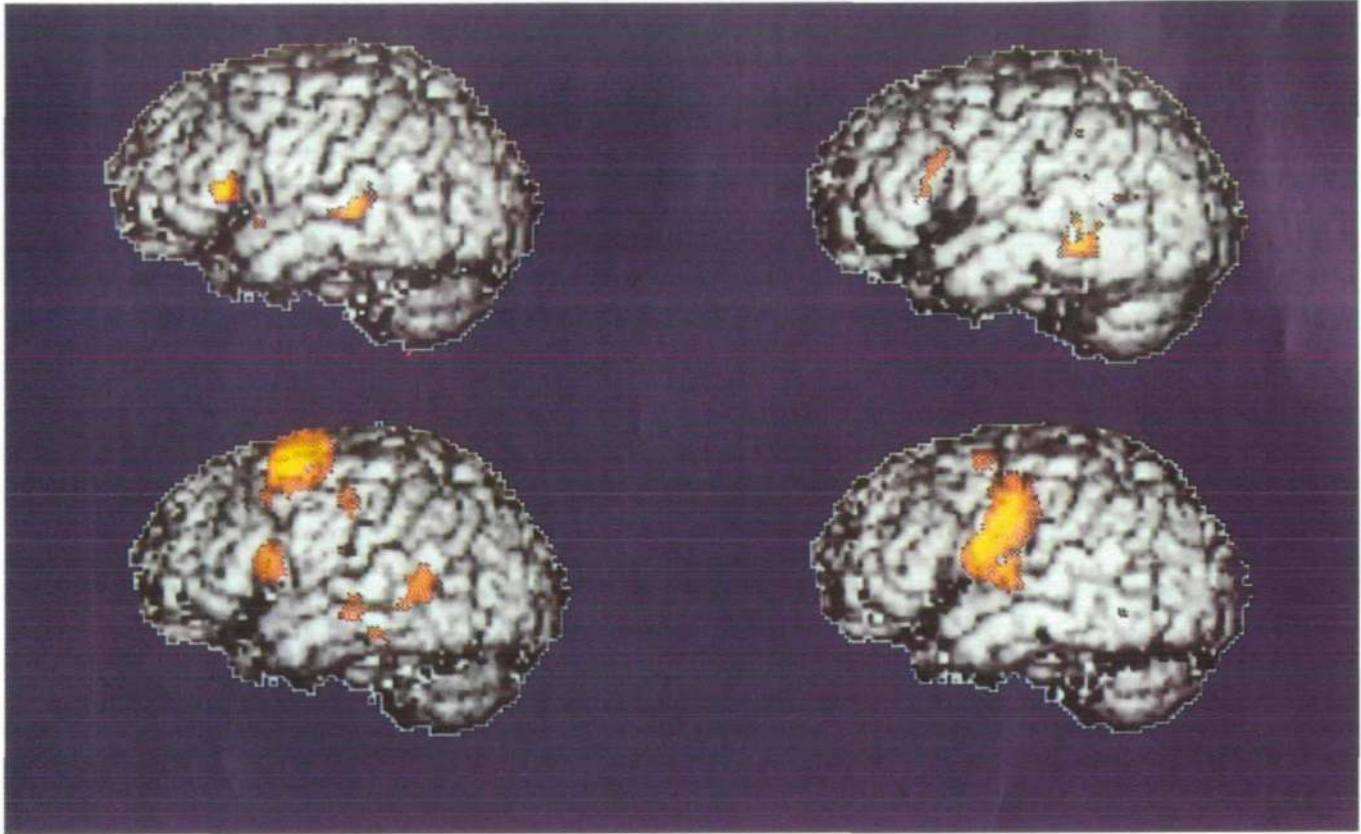


Fig. 4 The neural correlates of listening and repeating shown in red on grey models of the lateral surface of the left hemisphere. *Top left* is listening to words—listening to reversed words (Experiment 3); *top right* is the nonlinear response to hearing words (Experiment 5); *bottom left* is repeating words—saying ‘crime’ to reversed words (Experiment 2) and *bottom right* is repeating 40 w.p.m.—listening to 40 w.p.m. (Experiment 4).

Perception of own voice

In both mid-temporal lobes there was significantly greater activation during repeating 40 w.p.m. than during listening to 40 w.p.m., and the activation associated with repeating 20 w.p.m. matched that associated with listening to 40 w.p.m.. These results suggest that processing of self-monitored speech output, at least when repeating single words, is additive and equal to listening to words spoken by another. However, during reading aloud 40 w.p.m. the temporal regions were less active than when listening to 40 w.p.m.. This result may be explained by differences in attention; repeating and listening to words requires attention to be paid to the auditory percepts whereas during reading, attentional processes are directed towards the visual input, resulting in a lesser activation of auditory regions in response to one’s own voice. This explanation is supported by the results of a recent study by McGuire *et al.* (1996) that illustrate that when attention to the stimuli is controlled there is no difference in temporal activation when subjects read words aloud and hear the sound of their own voice through ear phones to when they read the same words silently and hear the voice of another speaker saying the words.

The demonstration that posterior temporal regions respond to the sound of one’s own voice is in accordance with

evidence that there is post-articulatory self-monitoring of speech output. For example, if a speaker’s voice is masked by noise he increases its intensity—the Lombard effect (Lane and Tranel, 1971)—but is less likely to detect certain types of speech errors (Lackner and Tuller, 1979). Levelt (1983, 1989) argues that the self-monitoring of speech relies on the same systems that are used for the analysis and comprehension of another’s speech, avoiding duplication of speech analysis systems. In contrast, single cell recordings in the primary auditory cortex of the squirrel monkey have shown that a population of cells are suppressed by phonation and that the responses to the animal’s spontaneous vocalizations and to the same vocalizations played back to the animal from tape differ (Muller-Preuss and Ploog, 1981). In the human, Creutzfeldt (1989), recording from left and right lateral temporal cortex during surgery, reported different responses of individual neurons to listening and speaking. Bilateral suppression of activity in a proportion of the neurons preceded vocalization by up to 100 ms and outlasted it for up to 1000 ms. PET activation studies record net activity with a spatial resolution of millions of synapses and a temporal resolution of many seconds and this study demonstrates that overall the effect of one’s own voice on the auditory and periauditory cortex is additive to perceiving the voice of another.

Areas involved in speech production

Articulation during repetition or reading aloud activates bilateral primary sensorimotor cortex (particularly on the left), bilateral anterior insular regions, the supplementary motor area, the basal ganglia, thalamic nuclei (particularly on the left) and the midline cerebellum. Activations of the left primary sensorimotor cortex, the supplementary motor area and the midline cerebellum were greater when different articulatory patterns were required for each word compared with when the same prespecified word was spoken on each occasion. The peaks of the primary sensorimotor activations were >20 mm above the intercommissural line, dorsal to the motor representations for lip, tongue and larynx. The observed activations probably represent voluntary control of respiration (Ramsay *et al.*, 1993), a necessary component of articulation. The reason why there were no separate activations observed for output to laryngeal, lingual and facial muscles in some of the contrasts relates to the proximity of sylvian sensorimotor cortex and auditory cortex. They lie within 10 mm of each other and simultaneous activation of both regions may show only one peak in the region with the higher Z score. By removing signals in the auditory cortex (for instance when repeating words was contrasted with listening to words), the sylvian sensorimotor activations are revealed (Table 4).

Posterior Broca's area (BA 44) or the adjacent precentral sulcus (BA 6) was activated by repetition and by reading aloud. These regions lie within a few millimetres of each other, and it will require single subject studies, with co-registration of the subject's PET data on to their MRI, to define accurately whether both of these structures are activated by repetition. The contrast between repeating and listening which reveals activity in BA 44, demonstrates that the peak of Broca's area activation for speech production lies 15–20 mm posterior to that associated with speech perception.

It is not possible to vary the rate of repetition independently of the rate of hearing the stimuli and of hearing one's own voice. This makes it difficult to isolate any component of the signal in Wernicke's area that might be contributing to speech production rather than speech perception. However, there was a clue from the response in the left posterior temporal gyrus. This is the region that is not sensitive to the rate of hearing words (see Fig. 2), but there was evidence that repetition did increase activation in this region when contrasted with listening to words. This suggests that, although net synaptic activity within this region reaches a maximum in response to low rates of input, further activity occurs with output.

Summary of main findings

The distributed systems for single word perception (bilateral temporal regions and anterior Broca's area) and single word production during repetition and reading (posterior Broca's area, bilateral anterior insular regions, bilateral sylvian and lateral primary sensorimotor cortex, supplementary motor

area, basal ganglia, thalamic nuclei and midline cerebellum) have been demonstrated in a series of three new experiments and reanalyses of two previously published studies. The main observations are as follows. (i) Perception of another's voice speaking single words is predominantly left lateralized once acoustic processing of complex non-speech sounds (in this instance, reversed words) has been accounted for. (ii) Perception of single words produces an activation in anterior Broca's area (probably BA 45) that is dependent on the rate of hearing words, being higher at slower rates. It is not apparent from this study whether the linguistic content of the auditory input or its tempo is responsible for this pattern of activation. (iii) Speech production involves posterior Broca's area (BA 44) and there may also be involvement of the immediately adjacent precentral sulcus (BA 6). (iv) There is a region in the left superior temporal gyrus where activity reaches a ceiling with low rates of hearing words but repetition produced additional activation in this region, suggesting a role for speech production as well as some aspect of speech perception. (v) The response in bilateral auditory association cortex to own voice during repetition is the same as when listening to another's voice from a prerecorded tape.

Acknowledgements

We wish to thank Graham Lewington and the radiochemists at the MRC Cyclotron Unit for their help in performing the studies, volunteers who underwent the scans, Stephanie Frackowiak for her help with image processing and Chris Freemantle for his help preparing the colour figure. The McDonnell Pew Program, grant number 91–23 funded C.J.P. The Stroke Association funded C.J.P. and C.J.M. The Wellcome Trust funded C.J.P., R.J.S.W., R.S.J.F. and K.F., and the Medical Research Council funded E.A.W..

References

- Bailey DL, Jones T, Watson JDG, Schnorr L, Frackowiak RSJ. Activation studies in 3D PET: evaluation of true signal gain. In: Uemura K, Lassen NA, Jones T, Kanno I, editors. Quantification of brain function. Amsterdam: Excerpta Medica, 1993: 341–7.
- Blumstein SE. The neurobiology of the sound structure of language. In Gazzaniga MS, editor. The cognitive neurosciences. Cambridge (MA): MIT Press, 1994: 915–29.
- Caplan D. Neurolinguistics and linguistic aphasiology: an introduction. Cambridge: Cambridge University Press, 1987.
- Creutzfeldt O, Ojemann G, Lettich E. Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects' own voice. *Exp Brain Res* 1989; 77: 476–89.
- Demonet J-F, Chollet F, Ramsay S, Cardebat D, Nespoulous J-D, Wise R, et al. The anatomy of phonological and semantic processing in normal subjects. *Brain* 1992; 115: 1753–68.
- Demonet J-F, Price C, Wise R, Frackowiak RSJ. A PET study of cognitive strategies in normal subjects during language tasks:

- influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain* 1994; 117: 671–82.
- Fox PT, Mintun MA. Non-invasive functional brain mapping by change-distribution analysis of averaged PET images of H215O tissue activity. *J Nucl Med* 1989; 30: 141–9.
- Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. Plastic transformation of PET image. *J Comput Assist Tomogr* 1991a; 15: 634–9.
- Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. Comparing functional (PET) images: The assessment of significant change. *J Cereb Blood Flow Metab* 1991b; 11: 690–9.
- Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC. Assessing the significance of focal activations using their spatial extent. *Hum Brain Map* 1994; 1: 210–20.
- Friston KJ, Ashburner J, Poline JB, Frith CD, Heather JD, Frackowiak RSJ. Spatial realignment and normalization of images. *Hum Brain Map* 1996. In press.
- Howard D, Patterson K, Wise R, Brown WD, Friston K, Weiller C, et al. The cortical localization of the lexicons: positron emission tomography evidence. *Brain* 1992; 115: 1769–82.
- Lackner JR, Tuller BH. Role of efference monitoring in the detection of self-produced speech errors. In sentence processing: psycholinguistic studies presented to Merrill Garrett, Cooper WE, Walker ECT, editors. Hillsdale (NJ): Lawrence Erlbaum, 1979: 281–94.
- Lammertsma AA, Cunningham VJ, Deiber MP, Heather JD, Bloomfield PM, Nutt J, et al. Combination of dynamic and integral methods for generating reproducible functional CBF images. *J Cerebr Blood Flow Metab* 1990; 10: 675–86.
- Lane H, Tranel B. The Lombard Sign and the role of hearing in speech. *J Speech Hear Res* 1971; 14: 677–709.
- Levelt WJM. Monitoring and self-repair in speech. *Cognition* 1983; 14: 41–104.
- Levelt WJM. Self-monitoring and self-repair. In: Levelt WJM. *Speaking: from intention to articulation*. Cambridge (MA): MIT Press 1989; 458–99.
- Mazziotta JC, Huang SC, Phelps ME, Carson RE, MacDonald NS, Mahoney K. A non-invasive positron computed tomography technique using oxygen-15 labeled water for the evaluation of neurobehavioral task batteries. *J Cereb Blood Flow Metab* 1985; 5: 70–8.
- McGuire P, Silberswieg D, Frith C. The functional anatomy of verbal self-monitoring. *Brain* 1996; 119: 907–17.
- Mesulam MM. Large-scale neurocognitive networks and distributed processing for attention, language, and memory. [Review]. *Ann Neurol* 1990; 28: 597–613.
- Muller-Preuss P, Ploog D. Inhibition of auditory cortical neurons during phonation. *Brain Res* 1981; 215: 61–76.
- Paulesu E, Frith CD, Frackowiak RSJ. The neural correlates of the verbal component of working memory [see comments]. *Nature* 1993; 362: 342–5. Comment in: *Nature* 1993; 363: 583–4.
- Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME. Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 1988; 331: 585–9.
- Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME. Positron emission tomographic studies of the processing of single words. *J Cognit Neurosci* 1989; 1: 153–70.
- Pinker S. *The language instinct: The new science of language and mind*. London: Allen Lane, 1994.
- Price C, Wise R, Ramsay S, Friston K, Howard D, Patterson K, et al. Regional response differences within the human auditory cortex when listening to words. *Neurosci Lett* 1992; 146: 179–82.
- Ramsay SC, Adams L, Murphy K, Corfield DR, Grootenok S, Bailey DL, et al. Regional cerebral blood flow during volitional expiration in man: a comparison with volitional inspiration. *J Physiol (Lond)* 1993; 461: 85–101.
- Robb RA, Hanson DP. A software system for interactive and quantitative visualization of multidimensional biomedical images. *Australas Phys Eng Sci Med* 1991; 14: 9–30.
- Silberswieg DA, Stern E, Frith CD, Cahill C, Schnorr L, Grootenok S, et al. Detection of thirty-second cognitive activations in single subjects with positron emission tomography: a new low-dose H215O regional cerebral blood flow three-dimensional imaging technique. *J Cereb Blood Flow Metab* 13: 617–29.
- Spinks TJ, Jones T, Gilardi MC, Heather JD. Physical performance of the latest generation of commercial positron scanner. *IEEE Trans Nucl Sci* 1988; 35: 721–5.
- Spinks TJ, Jones T, Bailey DL, Townsend DW, Grootenok S, Bloomfield PM, et al. Physical performance of a positron tomograph for brain imaging with retractable septa. *Phys Med Biol* 1992; 37: 1637–55.
- Talairach J, Tournoux P. *Co-planar stereotaxic atlas of the human brain*. Stuttgart: Thieme, 1988.
- Warburton E, Wise RJS, Price CJ, Weiller C, Hadar U, Ramsay S, et al. Noun and verb retrieval by normal subjects: studies with positron emission tomography. *Brain* 1996; 119: 159–79.
- 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.
- Woods RP, Cherry SR, Mazziotta JC. Rapid automated algorithm for aligning and reslicing PET images. *J Comput Assist Tomogr* 1992; 16: 620–33.
- Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for rCBF activation studies in human brain [see comments]. *J Cereb Blood Flow Metab* 1992; 12: 900–18. Comment in: *J Cereb Blood Flow Metab* 1993; 13: 1040–2.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. *Science* 1992; 256: 846–9.

Received August 14, 1995. Revised December 12, 1995. Accepted January 19, 1996

