A functional study of auditory verbal imagery

S. S. SHERGILL,[1] E. T. BULLMORE, M. J. BRAMMER, S. C. R. WILLIAMS,
R. M. MURRAY AND P. K. McGUIRE

From the Institute of Psychiatry and GKT School of Medicine and Dentistry, University of London

ABSTRACT

Background. We used functional MRI to examine the functional anatomy of inner speech and different forms of auditory verbal imagery (imagining speech) in normal volunteers. We hypothesized that generating inner speech and auditory verbal imagery would be associated with left inferior frontal activation, and that generating auditory verbal imagery would involve additional activation in the lateral temporal cortices.

Methods. Subjects were scanned, while performing inner speech and auditory verbal imagery tasks, using a 1.5 Tesla magnet.

Results. The generation of inner speech was associated with activation in the left inferior frontal/insula region, the left temporo-parietal cortex, right cerebellum and the supplementary motor area. Auditory verbal imagery in general, as indexed by the three imagery tasks combined, was associated with activation in the areas engaged during the inner speech task, plus the left precentral and superior temporal gyri (STG), and the right homologues of all these areas.

Conclusions. These results are consistent with the use of the 'articulatory loop' during both inner speech and auditory verbal imagery, and the greater engagement of verbal self-monitoring during auditory verbal imagery.

INTRODUCTION

Auditory verbal imagery refers to the process of imagining speech. This can occur naturally, e.g. when attempting to recall what someone said, but can also be deliberate, e.g. when consciously recalling the sound of someone's voice. Auditory imagery has been relatively understudied in comparison with the burgeoning psychological and functional neuroimaging literature on visual imagery, and there are even less data on auditory verbal imagery. Functional neuroimaging studies using single photon emission computerized tomography (SPECT) indicate that imagining non-verbal sounds, such as the barking of a dog, is associated with activation in the right temporal cortex (Goldenberg *et al*. 1991) and lesions in the right temporal cortex impair musical imagery (Zatorre & Halpern, 1993). Psychological studies (Reisberg *et al*. 1991;

Smith *et al*. 1995) suggest that the interaction between 'inner' speech and the 'inner ear' is crucial for effective auditory verbal imagery. The only neuroimaging study of auditory verbal imagery (AVI) to date used Positron Emission Tomography (PET) and found that imagining another person's speech, relative to generating inner speech in one's own voice, was associated with activation in the supplementary motor, anterior cingulate and precentral cortex, the fronto-parietal operculum, and the left superior temporal gyrus (McGuire *et al*. 1996).

Auditory verbal imagery is of particular interest because of its putative relationship with auditory verbal hallucinations. Cognitive models propose that auditory verbal hallucinations (AVH) are derived from inner speech that the patient has mis-identified as alien (Frith & Done, 1989). Auditory verbal imagery bears some resemblance to an AVH, in that the subject perceives the sound of another person's speech in the absence of an external stimulus. However, while normal inner speech seems to be

241

experienced mainly in the first person (e.g. 'I will …'; Hulbert *et al.* 1994), AVH in schizophrenia are usually in the second or third person e.g. 'You will …' or 'He will …' (Nyani & David, 1996). The reason for this apparent disparity between the grammatical forms of inner speech and AVH is unclear. One way of approaching this issue is to examine the relationship between the neural correlates of AVI in the first, second and third person. Inner speech and first person auditory verbal imagery are likely to differ from imagery in the second and third person because imagining another person's voice places greater demands on the generation and monitoring of inner speech (McGuire *et al.* 1996). Furthermore, there may also be differential engagement of visuospatial imagery; first person imagery involves only the subject, second person involves a speaker addressing the subject, while third person imagery involves a speaker referring to the subject as if in conversation with a third person.

In this study, we examined the neural correlates of auditory verbal imagery in normal subjects using functional magnetic resonance imaging (fMRI); this permitted a more powerful and sophisticated design than was previously possible using PET. Subjects performed four tasks: (*a*) the silent articulation of sentences; (*b*) imagining speech in their own voice (first person imagery); (*c*) imagining sentences spoken in another's voice, addressed to the subject (second person imagery); and (*d*) imagining sentences spoken by another about them (third person imagery).

We predicted the following.

1 All tasks would be associated with activation in the left inferior frontal gyrus/insula; corresponding to a common component of silent articulation (Paulesu *et al.* 1993; McGuire *et al.* 1996).

2 Auditory verbal imagery (in general) would be associated with activation in the lateral temporal cortex, corresponding to the monitoring of inner speech (McGuire *et al.* 1996).

3 Second and third person imagery would be associated with more activation in the left frontal and temporal cortex than imagining one's own speech, as imagining 'alien' speech places greater demands on the generation and monitoring of inner speech (McGuire *et al.* 1996).

4 Second and third person imagery would

also be associated with activation in the posterior parietal cortex, associated with the visuospatial imagery of the speaker in relation to another listener (Corbetta *et al.* 1993, 1995).

METHOD

Subjects

Eight male volunteers, right-handed according to Annett's (1970) scale, aged 26–37 years (mean age 33) participated in the study. They did not suffer from medical or psychiatric disorders and were not receiving medication, and had no family history of psychiatric disorder. Their mean IQ, estimated with the National Adult Reading Test (NART) (Nelson, 1981) was 114 (range 100–120). Potential subjects were assessed on their ability to perform the tasks (detailed below) outside the scanner. They rated their ability to perform each task on a visual analogue scale and were included if, after training, they consistently scored above 8/10 on this scale on all tasks. Subjects gave written informed consent to the study, which had been approved by the local hospital ethical committee.

Tasks performed during fMRI

All tasks required that subjects listened to single words presented every 3 s in the neutral voice that they had heard during training. The words consisted of nouns and adjectives that could readily complete a sentence of the form 'I like …', or 'I like being …'. The words were matched for familiarity, frequency of occurrence and imageability across conditions. Each task comprised an 'active' and a baseline condition. Before the baseline condition, subjects heard the prompt 'now listen', while the 'active' conditions were preceded by 'imagine'. Speech was presented via pneumatically driven headphones, incorporated within ear defenders, specifically designed for functional MRI (Quiet Muff 29 Earmuffs, Avotec, Jensen Beach, FL, USA). These reduced unattenuated noise from our gradient switching frequency for EPI (1 kHz) from 109 dB to approximately 68 dB. Attenuated noise at the second, third and fourth harmonics (2, 3, 4 kHz) was approximately 44–56 dB. Subjects had their eyes closed during scanning.

Four 'active' conditions were each compared with the same baseline condition. Each task

involved 10 alternating 30 s blocks of each condition, lasting a total of 5 min. Task order was counterbalanced across subjects.

Baseline

Subjects were instructed to listen to each word carefully.

Inner speech

On hearing each word, subjects were asked to silently articulate a sentence of the form 'I like …', or 'I like being …', ending in the presented word.

First person imagery

The instructions were identical to the inner speech condition, except that subjects had to imagine the sentence being spoken in their own voice.

Second person imagery

Subjects were asked to complete a sentence in the form 'You like …', or 'You like being …', ending in the presented word, and to imagine this spoken to them in the voice they had heard on the training tape.

Third person imagery

The instructions were identical to the second person imagery condition, except that the sentence was in the form 'He likes …', or 'He likes being …', as if spoken about the subject.

After completing each task, subjects rated their ability to perform it as instructed, using a visual analogue scale. In order to reduce potentially confounding effects of poor performance on activation, due to the distraction of being in the scanning environment, only data from subjects who provided ratings of greater than 5/10 on all tasks were analysed.

Image acquisition

Gradient-echo echoplanar MR images were acquired using a 1·5 Tesla GE Signa System (General Electric, Milwaukee, WI, USA) fitted with Advanced NMR hardware and software (ANMR, Woburn, MA, USA) at the Maudsley Hospital, London. A quadrature birdcage head coil was used for RF transmission and reception. In each of 14 non-contiguous planes parallel to the inter-commissural (AC–PC) plane, 100 T2*-weighted MR images depicting BOLD contrast (Ogawa *et al.* 1990) were acquired with TE = 40 ms, TR = 3000 ms, in-plane resolution = 3·1 mm, slice thickness = 7 mm, slice skip = 0·7 mm. Head movement was limited by foam padding within the head coil and a restraining band across the forehead. At the same session, a 43 slice, high-resolution inversion recovery echo-planar image of the whole brain was acquired in the AC–PC plane with TE = 73 ms, TI = 180 ms, TR = 16000 ms, in-plane resolution = 1·5 mm, slice thickness = 3 mm.

Generic brain activation mapping

Following correction of movement-related effects (Bullmore *et al.* 1999*a*), the power of periodic signal change at the (fundamental) ON–OFF frequency of stimulation was estimated by iterated least squares fitting a sinusoidal regression model to the motion-corrected time series at each voxel of all images. The fundamental power quotient (FPQ = fundamental power divided by its standard error) was estimated at each voxel and represented in a statistical map. Each observed fMRI time series was then randomly permuted 10 times, and FPQ re-estimated after each permutation. This resulted in 10 parametric maps (for each subject at each plane) of FPQ estimated under the null hypothesis that FPQ is not determined by experimental design (Bullmore *et al.* 1996). All parametric maps of FPQ were then registered in the standard space of Talairach & Tournoux (1988). This was achieved in two stages, using realignment algorithms similar to those previously used for movement correction. First, the set of FPQ maps observed in each subject was registered with that subject's high resolution EPI dataset; then registered and re-scaled relative to a Talairach template image. Identical transformations were applied to the randomized FPQ maps obtained for each subject. After spatial normalization, the observed and randomized FPQ maps from each subject were identically smoothed with a Gaussian filter (full width half maximum = 7 mm) to accommodate variability in gyral anatomy and error of voxel displacement during normalization. Generic activation was then decided by computing the median value of FPQ at each voxel of the observed parametric maps, and comparing it to a null distribution of median FPQ values computed from the randomized parametric

maps. If the observed median FPQ exceeded the critical value of randomized median FPQ, for a test of size $\alpha = 2 \cdot 5 \times 10^{-4}$, then that voxel was considered generically activated with probability of false positive activation $= \alpha$. Generically activated voxels were coloured and superimposed on the grey scale Talairach template, to create generic brain activation maps (GBAMs) (Brammer *et al.* 1997). The timing of the signal increase relative to the input function was ascertained and activated voxels with signal maximum during the active task were coloured red; activated voxels with signal maximum during baseline task were coloured blue.

The GBAMs for the imagery tasks were combined, taking the mean activation over all three imagery conditions *versus* baseline, to identify the consistent correlates of auditory verbal imagery. To estimate the between-task differences in the power of the physiological response to the experimental tasks, the following repeated measures analysis of variance (ANOVA) model was fitted at each intracerebral voxel in standard space:

$$\text{FPQ}_{i,j} = \beta_0 + \beta_1 \, \text{Task}_j + e_{i,j}.$$

Here, $\text{FPQ}_{i,j}$ denotes standardized power at the ith voxel in the jth member of the group; β_0 denotes the overall mean power, and $e_{i,j}$ is a residual quantity. Task is a factor coding for task (is or 1st and 2nd or 3rd). The null hypothesis of zero between-group difference in mean power of response was tested by comparing the coefficient to its non-parametrically ascertained null distribution. To do this the elements of Task were randomly permuted 10 times at each voxel, taking account of the repeated measures nature of the data (Edgington, 1980), β_1 was estimated after each permutation, and these estimates were pooled over all intracerebral voxels to sample the permutation distribution of β_1. Critical values of a 2-tailed test of size $P = 0 \cdot 01$ were the $100.(p/2)$th and $100.(1-p/2)$th percentiles of this distribution (Edgington, 1980; Bullmore *et al.* 1996, 1999*b*). Note that this relatively lenient probability threshold was used only to test a restricted search volume comprising those voxels that were generically activated by the task in one or both groups. Comparisons were made between: (*a*) inner speech and first person imagery (to examine the substrate of imagery), and (*b*)

first person imagery and combined second and third person imagery (to examine the correlates of 'self' *versus* 'alien' imagery).

RESULTS

Auditory verbal imagery ratings

Two individuals originally recruited for the study were unable to adequately perform the imagery tasks when inside the scanner, reporting that scanner noise made concentration difficult, and rating their performance on some tasks as less than 5/10. The data from these subjects was not analysed. The remaining six subjects reported that they were able to perform the tasks as instructed and scored above 5/10 on all tasks. They gave lower ratings for the tasks that involved imagining another person's voice, (second person imagery – mean score $7 \cdot 0$ (range $6 \cdot 5$–$9 \cdot 0$) and third person imagery – mean score $6 \cdot 5$ (range $6 \cdot 0$–$9 \cdot 0$)), than those that involved their own voice (inner speech – mean $9 \cdot 0$ (range $8 \cdot 5$–$10 \cdot 0$) and first person imagery – mean $8 \cdot 5$ (range $8 \cdot 0$–$9 \cdot 5$)); significantly different at $P < 0 \cdot 005$ ($t = 5 \cdot 6$, 95 % CI $1 \cdot 1$ to $3 \cdot 1$) (Fig. 1).

Imaging data

Inner speech

Compared to baseline, the inner speech condition was associated with left-sided activation in the inferior frontal gyrus, insula, the junction
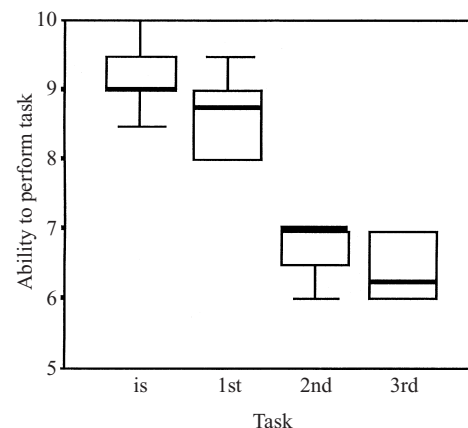

FIG. 1. The subjective ratings (scale 0–10), on the vertical axis, for each of the four conditions (on the horizontal axis); inner speech (is), first person (1st), second person (2nd) and third person (3rd). Boxplot displaying the mean value (bold line) with 95 % confidence intervals (shaded area) and range (whiskers).
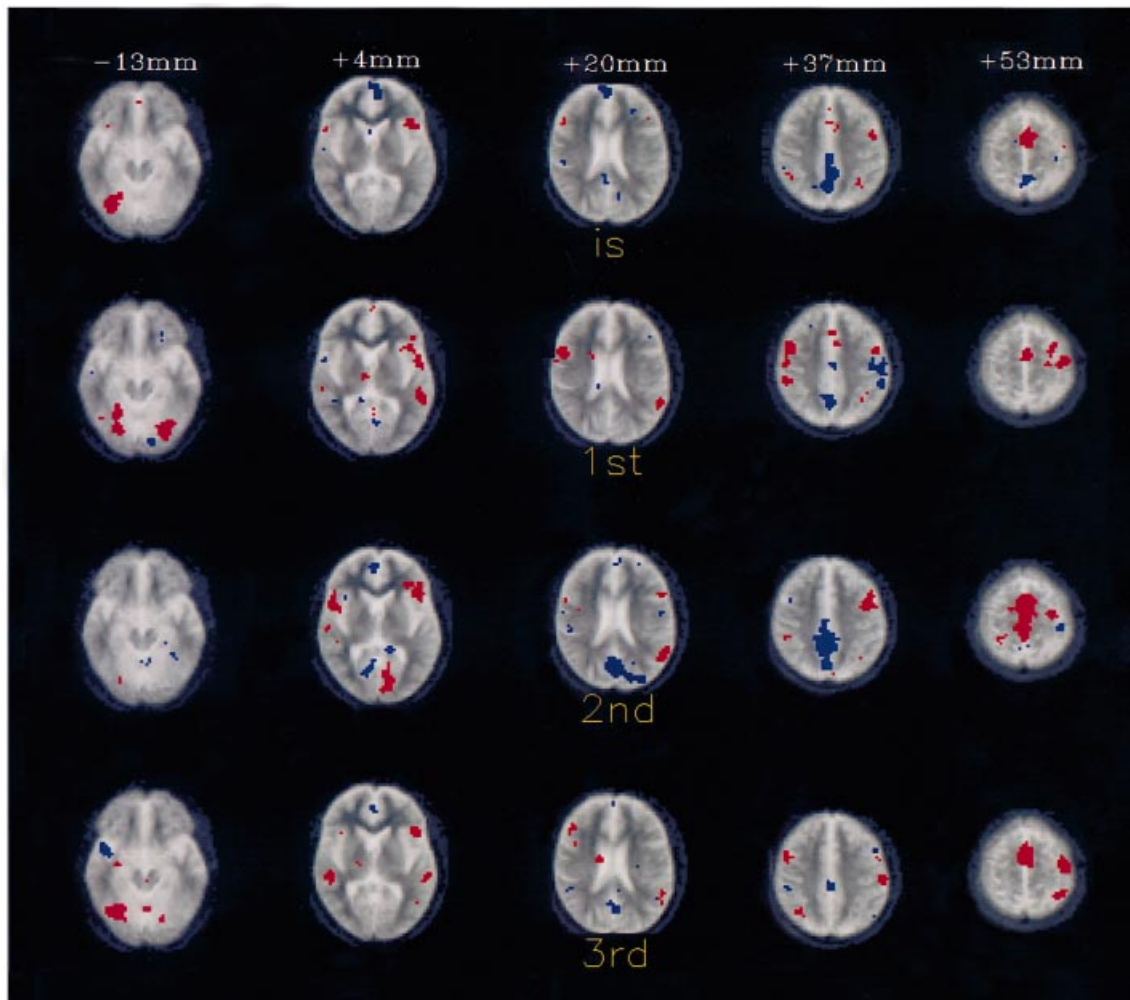
FIG. 2. Areas activated during: (*i*) inner speech relative to baseline; (*ii*) first person imagery relative to baseline; (*iii*) second person imagery relative to baseline; and (*iv*) third person imagery relative to baseline. Activation maps were rendered onto a template, generated from the coregistered structural MRIs from all six subjects, and correspond to Talairach space. Selected axial slices are displayed parallel to the anterior commissure–posterior commissure plane and the left side of the figure represents the right side of the subjects and vice versa. Areas shown in red depict clusters significantly activated during the first condition, while blue depicts regions activated during the baseline condition.

of the superior temporal gyrus (STG) and inferior parietal lobule, and the superior parietal lobule (Fig. 2). Activation was also evident in the supplementary motor area (SMA) and the right posterior cerebellar cortex. There were responses in phase with the baseline condition in the postcentral gyri, the medial prefrontal and parietal cortex and in the posterior cingulate gyrus (Table 1).

Imagining first person speech

Areas activated in phase with imagining one's own speech included the left insula, the precentral, postcentral and middle temporal gyri and the inferior parietal lobule (Fig. 2). Activation was also evident in the SMA, posterior cerebellar cortex bilaterally and the right inferior frontal gyrus, inferior parietal lobule, fusiform

Table 1. *Main foci of activation inner speech task*

Region of activation during	Voxels N	X	Y	Z	Max FPQ	P
Inner speech						
L STG/Inf parietal cortex (BA 22/39)	13	−52	−53	26	2·4	0·000005
R Posterior cerebellar cortex	29	26	−64	−18	2·2	0·000005
SMA (BA 6)	42	0	0	53	2·0	0·000015
L Sup parietal lobule (BA 7)	13	−26	−53	42	2·0	0·000015
L Inf frontal gyrus (BA 44)*	6	−46	11	26	1·9	0·00005
L Insula	7	−38	14	4	1·7	0·0005
Baseline						
Med frontal gyrus (BA 10)	24	0	56	4	2·2	0·000005
Post-cingulate gyrus (BA 31)	32	0	−42	37	2·1	0·000005
Med parietal cortex (BA 7)	26	3	−56	37	2·1	0·00001
R Post-central gyrus (BA 1, 2, 3)	6	52	−22	42	2·0	0·00002
L Post-central gyrus (BA 3)	4	−49	−14	42	1·9	0·00003

X, Y, Z, the coordinates of maximal response in the atlas of Talairach & Tournoux (1988).

Voxels N, number of suprathreshold voxels in cluster focused at this coordinate.

Max FPQ, Maximal value of the fundamental power quotient.

BA, Brodmann's Area; L, left; R, right; Inf, inferior; Sup, superior; STG, superior temporal gyrus; MTG, middle temporal gyrus; Ant, anterior; Med, medial; Mid, middle.

* Region of activation evident in the PET study of McGuire *et al.* (1996).

Table 2. *Main foci of activation, first person auditory verbal imagery*

Region	Voxels N	X	Y	Z	Max FPQ	P
Activation during imagining own speech						
SMA	21	3	−3	53	2·3	0·000005
L Inf parietal lobule (BA 40)	16	−49	−42	26	2·3	0·000005
R Posterior cerebellar cortex	10	26	−69	−13	2·3	0·000005
R Fusiform gyrus (BA 37)	22	29	−58	−7	2·2	0·000005
L Pre-central gyrus (BA 4)	19	−43	−11	48	2·2	0·000005
L Posterior cerebellar cortex	27	−23	−69	−13	2·1	0·00002
L Post-central gyrus (BA 1, 2, 3)	19	−43	−19	42	2·1	0·000015
L Mid-temporal gyrus (BA 21)	11	−52	−39	4	2·1	0·000005
R Inf parietal lobule (BA 40)	16	49	−31	42	2·0	0·00005
R Thalamus	14	12	−17	4	1·9	0·000075
R Inf frontal gyrus (BA 44)	10	49	3	20	1·8	0·00025
L Insula	6	−38	14	4	1·8	0·00025
Activation during baseline						
R Mid-frontal gyrus (BA 9)	7	49	8	42	2·3	0·000005
Medial frontal gyrus (BA 10)	17	3	53	9	2·1	0·000025
Post-cingulate gyrus/Med parietal cortex (BA 31/7)	13	6	−53	37	2·1	0·000005
Post-cingulate gyrus (BA 31)	12	3	−33	31	2·1	0·000005
L Lingual gyrus (BA 18)	7	−12	−83	−13	2·1	0·000015
Post-cingulate gyrus (BA 30)	20	3	−50	9	2·0	0·00005
R Hippocampus	11	23	−36	−2	2·0	0·00006

For explanation of abbreviations, see footnote to Table 1.

gyrus and thalamus. There were responses in phase with the baseline condition in the left lingual gyrus, the posterior cingulate/medial parietal cortex, the medial frontal gyrus and in the right middle frontal gyrus and hippocampus (Table 2).

Imagining second person speech

Regions activated in phase with imagining speech in the second person included the left inferior frontal, middle frontal and precentral gyri, the area around the junction of the left STG/inferior parietal lobule and the left posterior cerebellar cortex (Fig. 2). There was extensive activation of the SMA and responses in the right inferior frontal and superior temporal gyri. Responses in phase with the baseline condition were evident in the left cuneus and cerebellar vermis, the anterior cingulate gyrus, the area around the junction of the posterior cingulate gyrus and paracentral lobule, and in the medial parietal cortex (Table 3).

Table 3. *Main foci of activation, second person auditory verbal imagery*

Region	Voxels N	X	Y	Z	Max FPQ	P
Activation during second person speech						
SMA (BA 6)*	169	3	−3	48	3·0	0·000005
L STG/Inf parietal lobule (BA 22/40)*	34	−52	−44	26	2·3	0·000005
L Middle frontal gyrus (BA 9)	17	−43	11	31	2·1	0·000005
L Pre-central gyrus (BA 4)*	17	−43	−8	48	2·1	0·000005
L Cuneus (BA 17)	14	−6	−72	9	2·0	0·000005
R Inf frontal gyrus (BA 44)*	8	52	8	26	2·0	0·000005
L Inf frontal gyrus (BA 45)*	12	−38	19	4	1·9	0·000025
L Posterior cerebellar cortex	3	−9	−72	−7	1·8	0·000005
R Sup temporal gyrus (BA 22)	2	46	−25	9	1·6	0·00015
Activation during baseline						
Med parietal cortex (BA 7)	33	0	−61	42	2·8	0·000005
Ant cingulate gyrus (BA 32)	28	9	42	9	2·3	0·000005
Post-cingulate gyrus/Para-central lobule (BA 31/5)	54	6	−31	42	2·2	0·000005
Post-cingulate gyrus/Para-central lobule (BA 31/5)	30	0	−42	37	2·1	0·000005
L Cuneus (BA 19)	20	−26	−78	20	2·0	0·000005
L Cerebellar vermis	8	−9	−75	−18	2·0	0·000005

For explanation of abbreviations, see footnote to Table 1.
* Region of activation evident in the PET study of McGuire *et al.* (1996).

Table 4. *Main foci of activation, third person auditory verbal imagery*

Region	Voxels N	X	Y	Z	Max FPQ	P
Activation during third person imagery						
L Pre-central gyrus (BA 4)	60	−40	−14	53	2·5	0·000005
SMA (BA 6)	46	3	−3	53	2·4	0·000005
L Post-central gyrus (BA 1, 2, 3)	23	−52	−19	31	2·3	0·000005
R STG (BA 22)	28	49	−36	9	2·1	0·00001
R Posterior cerebellar cortex	24	32	−64	−13	2·1	0·000005
L Inf parietal lobule (BA 40)	18	−52	−44	26	2·1	0·000005
L STG (BA 42)	17	−58	−31	9	2·1	0·000015
L Insula	9	−40	17	4	2·1	0·00001
R Pre-central gyrus/Inf frontal gyrus (BA 6/44)	19	55	6	31	2·0	0·000075
R STG (BA 38)	8	46	3	−13	1·9	0·000175
Activation during baseline						
Posterior cingulate gyrus/medial parietal cortex (BA 31/7)	17	−3	−67	26	2·1	0·000015
Rostralant cingulate cortex (BA 32)	12	3	44	15	2·1	0·00002
Post-cingulate gyrus (BA 23)	12	3	−42	26	2·0	0·000075

For explanation of abbreviations, see footnote to Table 1.

Imagining third person speech

Imagining speech in the third person was associated with left sided activation in the insula, pre- and post-central gyri, STG and the inferior parietal lobule (Fig. 2). Further responses were evident in the SMA, the area around the junction of the right inferior and middle frontal gyri, the right STG and the right posterior cerebellar cortex. Responses in phase with the baseline condition were apparent in the anterior and posterior cingulate gyri, and around the junction of the posterior cingulate gyrus and medial parietal cortex (Table 4).

Auditory verbal imagery (*first, second and third person imagery combined*)

Activation in phase with the combined imagery conditions was evident in the left inferior frontal and precentral gyri, and around the junction of the left STG with the inferior parietal lobule (Fig. 3). Midline activation was evident in the SMA, and there were right-sided responses in the inferior frontal, precentral and superior temporal gyri, the inferior parietal lobule and the posterior cerebellar cortex. Responses in phase with the baseline task were evident in the medial prefrontal cortex, around the junction of
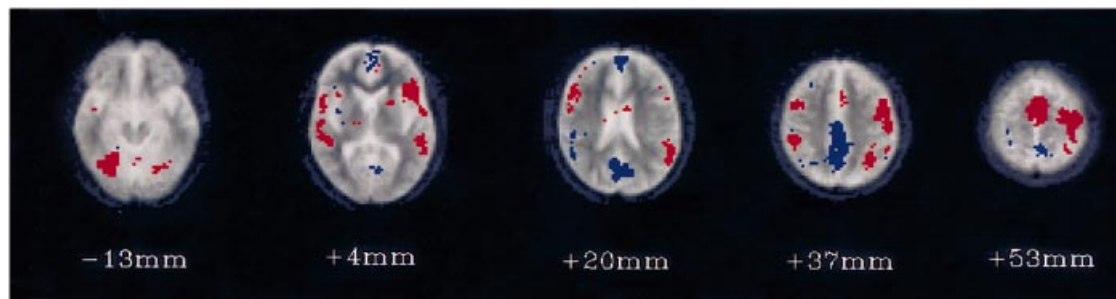
FIG. 3. Areas activated during auditory verbal imagery (first, second and third person imagery combined, relative to baseline). Activation maps were rendered onto a template, generated from the coregistered structural MRIs from all six subjects, and correspond to Talairach space. Selected axial slices are displayed parallel to the anterior commisure–posterior commisure plane and the left side of the figure represents the right side of the subjects and vice versa. Areas shown in red depict clusters significantly activated during the first condition, while blue depicts regions activated during the baseline condition.

Table 5. *Main foci of activation, during combined imagining speech*

Region	Voxels N	X	Y	Z	Max FPQ	P
Activation during combined imagining speech						
SMA (BA 6)	87	3	−3	53	2·5	0·000005
L STG/Inf parietal lobule (BA 22/40)	36	−49	−42	26	2·3	0·000005
L Inf frontal gyrus (BA 45)	14	−43	17	4	2·1	0·00001
L Pre-central gyrus (BA 4)	41	−40	−8	48	2·0	0·000005
R Inf frontal gyrus (BA 44)	20	52	8	26	1·7	0·000005
R Cerebellar cortex	18	26	−58	−13	1·7	0·000005
R STG (BA 22)	15	46	−36	4	1·7	0·000005
L Pre-central gyrus (BA 6)	14	−40	−3	37	1·7	0·000005
L Inferior parietal lobule (BA 40)	12	−38	−39	48	1·7	0·000005
R Inf parietal lobule (BA 40)	11	46	−31	37	1·6	0·00002
R Pre-central gyrus (BA 6)	8	43	3	37	1·6	0·00001
Activation during baseline						
R Inf parietal lobule (BA 40)	15	58	−28	26	1·8	0·000005
Medial frontal lobe (BA 10)	26	6	50	−2	1·8	0·000005
Post-cingulate gyrus (BA 23)	31	3	−33	42	1·8	0·000005
Posterior cingulate gyrus/Medial parietal cortex (BA 31/7)	52	0	−64	20	1·7	0·000005

For explanation of abbreviations, see footnote to Table 1.

the posterior cingulate gyrus and in the medial parietal cortex, and the right inferior parietal lobule (Table 5).

Differences between tasks

Only differences significant at $P < 0.005$ are reported.

First person imagery versus inner speech

There was a greater response during imagining one's own voice, relative to inner speech, in the left insula, precentral gyrus and lingual gyrus and bilaterally in the middle temporal gyri and posterior cerebellar cortex (Fig. 4). Right-sided changes were evident in the middle frontal gyrus, inferior parietal lobule, fusiform gyrus, the hippocampus and the thalamus. During the inner speech condition there was a greater response in the SMA, during the active task, and in the medial frontal gyrus, anterior and posterior cingulate gyri and the medial parietal lobule in phase with the baseline condition in the frontal pole, right middle frontal gyrus, left lingual and posterior cingulate gyri.

Second and third person imagery combined versus first person imagery

Imagery of another's (second and third person imagery combined) voice was associated with greater activation in the SMA, the left precentral and middle temporal gyri and inferior parietal lobule, and right superior temporal gyrus and posterior cerebellar cortex (Fig. 4). Second and third person imagery combined was also
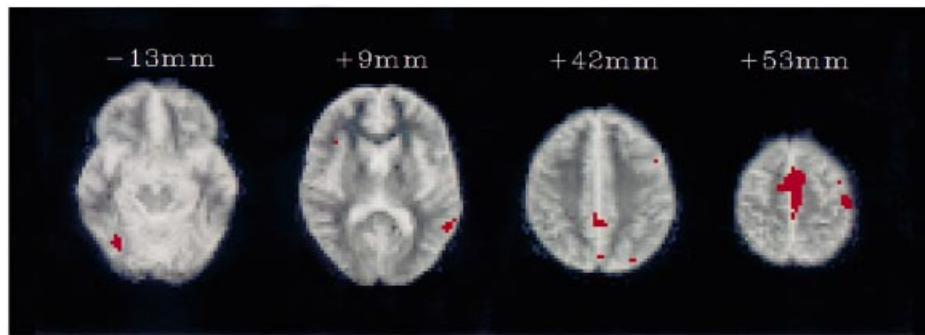
FIG. 4. Areas activated during combined second and third person imagery relative to first person. Selected axial slices are displayed parallel to the anterior commisure–posterior commisure plane and the left side of the figure represents the right side of the subjects and vice versa. Areas shown in red depict clusters significantly activated during the first condition.

Table 6. ***Combined second and third person imagery*** versus *first person imagery*

Region	Voxels N	X	Y	Z	Max β_1	P
SMA (BA6)	73	3	11	53	0·5	0·000005
L Pre-central gyrus (BA 4)	17	−40	−11	53	0·5	0·000005
R Posterior cerebellar cortex	9	35	−58	−13	0·5	0·000005
L MTG (BA 21)	7	−58	−44	9	0·5	0·000005
L Inf parietal lobule (BA 40)	7	−55	−36	26	0·5	0·000005
R STG (BA 22)	4	49	−25	4	0·5	0·000005
Medial parietal lobule (BA 7)*	9	−3	−58	42	0·7	0·000005
Post-cingulate gyrus (BA 31)*	16	6	−33	37	0·6	0·000005

For explanation of abbreviations, see footnote to Table 1.
Bold type indicates: greater activation during first task.
* Greater activation in phase with baseline/('decrease') during the active condition.

associated with greater responses in the medial parietal lobule and posterior cingulate gyrus during the baseline condition (Table 6).

DISCUSSION

Methodology

The aim of this study was to use fMRI to identify the neural correlates of inner speech and different forms of auditory verbal imagery. The inner speech and imagery tasks required subjects to generate and silently articulate/imagine sentences from cue words. The baseline condition (listening to the cue words) was designed to eliminate activation related to processing the cue, isolating responses specific to inner speech and auditory verbal imagery. The use of a stereotyped sentence prevented variation in the form of sentence construction beyond that specified by the task instructions.

Although all the original subjects were able to perform the tasks when outside the scanner, two proved unable to do so once lying inside it; their

fMRI data was therefore not analysed. The remaining subjects reported that the scanner noise (which was reduced to circa 60 db by ear defenders) was occasionally a distraction but did not interfere with task execution. Although scanner noise can also influence the BOLD response in lateral temporal cortex (Bandettini *et al.* 1998; Amaro *et al.* 1999) significant, task-related, temporal cortical responses were still evident in this study, presumably because confounding effects of background noise were reduced by its constancy across conditions. Previous fMRI studies using the same magnet have also detected temporal cortical activation both in response to external speech (Woodruff *et al.* 1997) and during tasks engaging covert verbal processing (Calvert *et al.* 1997). Ideally, functional imaging studies of cognitive tasks should involve the measurement of behavioural performance 'on-line', to assess how well subjects are carrying them out during scanning. However, auditory verbal imagery is a subjective phenomenon, and as such is not easily amenable

to objective monitoring of performance. We thus sought to minimize the influence of variation in performance by training subjects on the tasks prior to scanning, obtaining ratings of their performance immediately after each task, and excluding those who reported marked difficulties with task execution.

Verbal generation and silent articulation

As predicted, the inner speech condition, which entailed the silent generation and articulation of speech, was associated with activation in the left inferior frontal cortex and insula. However, there was also activation in several regions that were not activated in our previous PET study of auditory verbal imagery (McGuire *et al.* 1996), such as the SMA, the left superior temporal/inferior parietal region, and the right posterior cerebellar cortex. This probably reflects the greater statistical power of the present study which, because it employed fMRI, involved 400 images per subject as opposed to 12. While the inner speech condition involved semantic and grammatical processing that could have contributed to this pattern of activation, the use of a stereotyped sentence stem meant that the main demands were on covert generation/articulation and verbal working memory. Moreover, similar areas have previously been activated with other tasks that have engaged verbal working memory (Paulesu *et al.* 1993) and silent articulation implicitly (Warburton *et al.* 1996; Curtis *et al.* 1998). The role of the right cerebellum in language and verbal working memory tasks has become more apparent with the increased use of functional neuroimaging, reviewed by Desmond & Fiez (1998); they suggest that the cerebellum displays activation both during cognitive manipulations requiring verbal working memory and those involving articulatory rehearsal. The same set of regions were also engaged during first, second and third person imagery, consistent with the notion that inner speech and auditory verbal imagery share a common component of verbal working memory and covert generation/articulation (Smith *et al.* 1995; McGuire *et al.* 1996).

Auditory verbal imagery

The auditory verbal imagery tasks probably placed greater demands on both covert generation and articulation and auditory attentional engagement than the inner speech task, because the subject had to reproduce an internal representation of speech with a particular sound (McGuire *et al.* 1996). This would be consistent with our subjects rating the imagery tasks as subjectively more difficult than the inner speech task. Imagining speech is thought to entail retrieval of the voice to be imagined from memory and the internal reproduction and 'inspection' of the auditory image (Reisberg *et al.* 1991). Auditory verbal imagery in general, as indexed by the three imagery tasks combined, was associated with activation in the set of areas engaged during the inner speech task (left inferior frontal and parietal cortex and SMA), plus the left precentral and superior temporal gyri, and the right homologues of all these areas. Similar regions were activated, albeit less extensively, in the comparison between first person imagery and inner speech. We predicted the lateral temporal activation because this region is implicated in verbal monitoring, which is greater during auditory verbal imagery than inner speech (Wise *et al.* 1991; Zatorre & Halpern, 1993; McGuire *et al.* 1996); while the engagement of the left inferior frontal gyrus/insula and precentral cortex during auditory verbal imagery was predicted because imagining speech places increased demands on silent articulation (McGuire *et al.* 1996; Paulescu, 1993; Price *et al.* 1996). The additional engagement of the right-sided analogues of these areas suggests that they may play a role in retrieving and processing the prosody of the imagined speech, as they are activated when comparing melodies with matched non-melodic sequences, making pitch judgements on syllables (Zatorre & Sampson, 1991; Zatorre *et al.* 1992), assessing emotional prosody in spoken sentences (George *et al.* 1996) and implicated in imagery of familiar tunes (Halpern & Zatorre, 1999). The latter PET study of musical imagery also suggested that right frontal lobe activation was associated with retrieval from musical semantic memory, and confirmed the involvement of the SMA in image generation, proposing that it may be related to motor code in this process. These regions have also been associated with deficits in processing verbal prosody in lesion studies (Ross & Mesulam, 1979; Pell, 1998). However, an alternative, more general, explanation could be that increased task demands led to the recruitment of

contralateral homologous regions, as evident in a recent fMRI study of mental rotation (Carpenter *et al.* 1999).

Imagining alien speech

The combined second and third person imagery, relative to first person imagery, was associated with activation in the left premotor, middle temporal and inferior parietal cortex, and in the SMA. The greater engagement of the premotor, inferior parietal and temporal cortex are consistent with the notion that imagining another's voice places more demands on covert articulation, engagement of auditory attention and on verbal monitoring, than imagining one's own (McGuire *et al.* 1996).

There was no evidence to support the hypothesis of greater activation of cortical visuospatial areas during imagery of 'alien' auditory imagery. The most likely reason for this failure is that the instructions did not require subjects to evoke any visual imagery during these tasks.

Baseline responses

Although there were differences in the extent to which they were engaged, the medial parietal and posterior cingulate cortex showed responses during the baseline of all the tasks. These changes may reflect a net shift of attention from visuospatial to language processing during the active condition such that resources are diverted away from areas irrelevant to the active task (McGuire *et al.* 1996). Analogous changes, in the opposite direction, have been reported in the auditory cortex, during visual paradigms (Haxby *et al.* 1994). A similar mechanism may account for the baseline responses we observed in the medial frontal cortex, as increased baseline activity in this region has also been noted during auditory verbal memory tasks (Grasby *et al.* 1994).

Auditory verbal hallucinations

The areas engaged during inner speech and auditory verbal imagery in this study, particularly the inferior frontal/insular, lateral temporal and inferior parietal cortex, are similar to those identified in functional imaging studies of auditory verbal hallucinations (McGuire *et al.* 1993, 1996; Silbersweig *et al.* 1995; Shergill *et al.* 2000*a*). This is consistent with the notion that verbal hallucinations are derived from a lack of awareness of the subject's normal inner speech (Frith, 1992). However, patients with hallucinations are not usually conscious of generating the experience, whereas when volunteers imagine other people's speech they have no difficulty in recognizing this as self-generated. In the present study, both the SMA and cerebellum were strongly activated during the inner speech and auditory verbal imagery tasks, particularly when subjects imagined 'alien' speech, but it appears to be relatively weakly engaged during auditory verbal hallucinations (Silbersweig *et al.* 1995; Dierks *et al.* 1999; Shergill *et al.* 2000*a*). The SMA is thought to be the source of the *Bereitschaft* potential, which precedes the awareness of willed action (Deeke *et al.* 1969; Deeke, 1987), and lesions in this region are associated with alien limb syndrome, in which the patient seems to lose awareness that his movements are self-generated (Gasquoine, 1993). Similarly, both lesion and neuroimaging studies suggest that the cerebellum normally acts as a 'comparator' in both motor (Blakemore *et al.* 1998) and verbal tasks (Desmond *et al.* 1997), comparing intended with actual performance and modulating cerebral cortical activity appropriately (Andreasen *et al.* 1998). These observations suggest activation in the SMA or the cerebellum may be a correlate of the subject being aware that an auditory verbal percept is self-generated. Conversely, a lack of SMA or cerebellar activation, as seems to occur, during auditory hallucinations might be related to a loss of this awareness. This is consistent with a case report of a patient with right lateral cerebellar damage, which resulted in an inability to correct verbal errors, of which the patient was often unaware (Fiez *et al.* 1992). Similarly, a functional neuroimaging study of auditory imagery in patients with schizophrenia, in remission, demonstrated hypoactivity within the cerebellum, as well as other regions (Shergill *et al.* 2000*b*).

REFERENCES

Amaro, E., Brammer, M. J., Williams, S. C. R., Andrew, C., Curtis, V., Ahmad, F., Murray, R. M. & McGuire, P. K. (1999). Event-related fMRI without scanner acoustic noise. *Schizophrenia Research* **36**, 217.

Andreasen, N. C., Paradiso, S. & O'Leary, D. S. (1998). 'Cognitive dysmetria' as an integrative theory of schizophrenia: a dysfunction in cortical-subcortical-cerebellar circuitry? *Schizophrenia Bulletin* **24**, 203–218.

Annett, M. A. (1970). A classification of hand preference by association analysis. *British Journal of Psychology* **61**, 303–321.

Bandettini, P. A., Jesmanowicz, A., Van Kylen, J., Birn, R. M. & Hyde, J. S. (1998). Functional MRI of brain activation induced by scanner acoustic noise. *Magnetic Resonance in Medicine* **39**, 410–416.

Blakemore, S. J., Wolpert, D. M. & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience* **1**, 635–640.

Brammer, M. J., Bullmore, E. T., Simmons, A., Williams, S. C. R., Grasby, P. M., Howard, R. J., Woodruff, P. W. & Rabe-Hesketh, S. (1997). Generic brain activation mapping in fMRI: a non-parametric approach. *Magnetic Resonance Imaging* **15**, 763–770.

Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Rabe-Hesketh, S., Janot, N., David, A. S., Mellers, J. D. C., Howard, R. & Sham, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine* **35**, 261–277.

Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Curtis, V., Morris, R. G., Williams, S. C. R., Sharma, T. & McGuire, P. K. (1999a). Some methods for diagnosis and treatment of stimulus correlated motion in fMRI studies of brain activation. *Human Brain Mapping* **7**, 38–48.

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E. & Brammer, M. J. (1999b). Global, voxel and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging* **18**, 32–42.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D. & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science* **276**, 593–596.

Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W. & Thulborn, K. (1999). Graded functional activation in the visuospatial system with the amount of task demand. *Journal of Cognitive Neuroscience* **11**, 9–24.

Corbetta, M., Miezin, F. M., Shulman, G. L. & Petersen, S. E. (1993). A PET study of visuospatial attention. *Journal of Neuroscience* **13**, 1202–1226.

Corbetta, M., Shulman, G. L., Miezin, F. M. & Petersen, S. E. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science* **270**, 802–805.

Curtis, V. A., Bullmore, E. T., Brammer, M. J., Wright, I. C., Williams, S. C. R., Morris, R. G., Sharma, T. S., Murray, R. M. & McGuire, P. K. (1998). Attenuated frontal activation during a verbal fluency task in patients with schizophrenia. *American Journal of Psychiatry* **155**, 1056–1063.

Deeke, L. (1987). Bereitschaftspotential as an indicator of movement preparation in supplementary motor area and motor cortex. In *Motor Areas of the Cerebral Cortex* (ed. R. Porter), pp. 231–250. Wiley: Chichester.

Deeke, L., Scheid, P. & Kornhuber, H. H. (1969). Distribution of readiness potential, premotion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements. *Experimental Brain Research* **7**, 158–168.

Desmond, J. E., Gabrieli, J. D. E., Wagner, A. D., Ginier, B. L. & Glover, G. H. (1997). Lobular patterns of cerebellar activations in verbal working memory and finger tapping tasks as revealed by functional MRI. *Journal of Neuroscience* **17**, 9675–9685.

Dierks, T., Linden, D. E. J., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H. & Singer, W. (1999). Activation of Heschl's gyrus during auditory hallucinations. *Neuron* **22**, 615–621.

Desmond, J. E. & Fiez, J. A. (1998). Neuroimaging studies of the cerebellum: language, learning and memory. *Trends in Cognitive Sciences* **2**, 355–361.

Edgington, E. S. (1980). *Randomisation Tests*. Marcel Dekker: New York.

Fiez, J. A., Petersen, S. E., Cheney, M. K. & Raichle, M. E. (1992). Impaired non-motor learning and error detection associated with cerebellar damage. A single case study. *Brain* **115**, 155–178.

Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Lawrence Erlbaum Associates Ltd: New York.

Frith, C. D. & Done, D. J. (1989). Towards a neuropsychology of schizophrenia. *British Journal of Psychiatry* **153**, 437–443.

Gasquoine, P. G. (1993). Alien hand sign. *Journal of Clinical and Experimental Neuropsychology* **5**, 653–666.

George, M. S., Parekh, P. I., Rosinsky, N., Ketter, T. A., Kimbrell, T. A., Heilman, K. M., Herscovitch, P. & Post, R. M. (1996). Understanding emotional prosody activates right hemisphere regions. *Archives of Neurology* **53**, 665–670.

Goldenberg, G., Podreka, I., Steiner, M., Franzen, P. & Deecke, L. (1991). Contributions of occipital land temporal brain regions visual and acoustic imagery – a SPECT study. *Neuropsychologia* **29**, 695–702.

Grasby, P. M., Frith, C. D., Friston, K. J., Simpson, J., Fletcher, P. C., Frackowiak, R. S. J. & Dolan, R. J. (1994). A graded task approach to the functional mapping of brain areas implicated in auditory-verbal memory. *Brain* **117**, 1271–1282.

Halpern, A. R. & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex* **9**, 697–704.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P. & Grady, C. L. (1994). The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience* **14**, 6336–6353.

Hulbert, R. T., Happé, F. & Frith, U. (1994). Sampling the form of inner experience in three adult with Asperger's syndrome. *Psychological Medicine* **24**, 385–395.

McGuire, P. K., Shah, G. M. S. & Murray, R. M. (1993). Increased blood flow in Broca's area during auditory hallucinations in schizophrenia. *Lancet* **342**, 703–706.

McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., David, A. S., Frackowiak, R. J. S. & Frith, C. D. (1995). Abnormal monitoring of inner speech: a physiological basis for auditory hallucinations. *Lancet* **346**, 596–600.

McGuire, P. K., Silbersweig, D. A. & Frith, C. D. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine* **26**, 29–38.

Nelson, H. E. (1991). *National Adult Reading Test* (*NART*). NFER-Nelson: Windsor, Berks.

Nayani, T. H. & David, A. S. (1996). The auditory hallucination: a phenomenological survey. *Psychological Medicine* **26**, 177–189.

Ogawa, S., Lee, T. M., Kay, A. R. & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings National Academy of Science* (*USA*) **3**, 9868–9872.

Paulesu, E., Frith, C. D. & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature* **362**, 342–345.

Pell, M. D. (1998). Recognition of prosody following unilateral brain lesion: Influence of functional and structural attributes of prosodic contours. *Neuropsychologia* **36**, 701–715.

Price, C. J., Wise, R. J. S., Warburton, E., Moore, C. J., Howard, D., Patterson, K., Frackowiak, R. J. S. & Friston, K. J. (1996). Hearing and saying: the functional neuroanatomy of auditory word processing. *Brain* **119**, 919–931.

Reisberg, D., Wilson, M. & Smith, J. D. (1991). Auditory imagery and inner speech. In *Mental Images in Human Cognition* (ed. R. Logie and M. Denis), pp. 59–81. Elsevier: Amsterdam.

Ross, E. D. & Mesulam, M. M. (1979). Dominant language functions of the right hemisphere? Prosody and emotional gesturing. *Archives of Neurology* **36**, 144–149.

Shergill, S. S., Brammer, M. J., Williams, S., Murray, R. M. & McGuire, P. K. (2000a). Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. *Archives of General Psychiatry* **57**, 1033–1038.

Shergill, S. S., Bullmore, E. T., Simmons, A., Murray, R. M. & McGuire, P. K. (2000b). The functional anatomy of auditory

verbal imagery in patients with auditory hallucinations. *American Journal of Psychiatry* **157**, 1691–1693.

Silbersweig, D. A., Stern, E., Frith, C. D., Cahill, C., Holmes, A., Grootoonk, S., Seaward, J., McKenna, P., Chua, S. E. & Schnorr, L. (1995). A functional neuroanatomy of hallucinations in schizophrenia. *Nature* **378**, 176–179.

Smith, J. D., Wilson, M. & Reisberg, D. (1995). The role of subvocalisation in auditory imagery. *Neuropsychologia* **33**, 1433–1454.

Talairach, J. & Tournoux, P. (1988). *A Coplanar Stereotactic Atlas of the Human Brain*. Thieme Verlag: Stuttgart.

Warburton, E., Wise, R. J. S., Price, C. J., Weller, C., Hadar, U., Ramsay, S. & Frackowiak, R. J. S. (1996). Noun and verb retrieval by normal subjects. Studies with PET. *Brain* **119**, 159–179.

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.

Woodruff, P. W., Wright, I. C., Shuriquie, N., Russouw, H., Rushe, T., Howard, R. J., Graves, M., Bullmore, E. T. & Murray, R. M. (1997). Auditory hallucinations and the temporal cortical response to speech in schizophrenia: a functional magnetic resonance imaging study. *American Journal of Psychiatry* **154**, 1676–1682.

Zatorre, R. J. & Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia* **31**, 221–232.

Zatorre, R. J. & Samson, S. (1991). Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain* **114**, 2403–2417.

Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science* **256**, 846–849.