Dissociable Functional Cortical Topographies for Working Memory Maintenance of Voice Identity and Location

Pia Rämä[1,2], Amy Poremba[3], Joseph B. Sala[1], Lydia Yee[1], Megan Malloy[4], Mortimer Mishkin[4] and Susan M. Courtney[1,5]

[1]Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA, [2]Cognitive Brain Research Unit, Department of Psychology, University of Helsinki and Helsinki Brain Research Centre, Finland, [3]Department of Psychology, University of Iowa, Iowa City, IA, USA, [4]Laboratory of Neuropsychology, NIMH, Bethesda, MD, USA and [5]F.M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, MD, USA

In order to ascertain whether the neural system for auditory working memory exhibits a functional dissociation for spatial and nonspatial information, we used functional magnetic resonance imaging and a single set of auditory stimuli to study working memory for the location and identity of human voices. The subjects performed a delayed recognition task for human voices and voice locations and an auditory sensorimotor control task. Several temporal, parietal, and frontal areas were activated by both memory tasks in comparison with the control task. However, during the delay periods, activation was greater for the location than for the voice identity task in dorsal prefrontal (SFS/PreCG) and parietal regions and, conversely, greater for voices than locations in ventral prefrontal cortex and the anterior portion of the insula. This preferential response to the voice identity task in ventral prefrontal cortex continued during the recognition test period, but the double dissociation was observed only during maintenance, not during encoding or recognition. Together, the present findings suggest that, during auditory working memory, maintenance of spatial and nonspatial information modulates activity preferentially in a dorsal and a ventral auditory pathway, respectively. Furthermore, the magnitude of this dissociation seems to be dependent on the cognitive operations required at different times during task performance.

Keywords: auditory system, functional magnetic resonance imaging, nonspatial memory, prefrontal cortex, spatial memory

Introduction

Recently, it has been proposed that the auditory system is organized into two domain-specific, spatial and nonspatial, processing streams, similar to that seen in the visual system (Ungerleider and Mishkin, 1982; Ungerleider and Haxby, 1994). Different regions of the auditory cortex have been shown to be differentially responsive to spatial and spectral features of auditory stimulation (Romanski et al., 1999; Tian et al., 2001; for review see Rauschecker and Tian, 2000). The caudolateral (CL) region exhibits greater selectivity for spatial features of stimuli whereas neurons in the anterolateral (AL) region show greater selectivity for monkey vocalization. These regions in the auditory belt cortex have been shown to project to distinct regions of the temporal, parietal, and prefrontal cortices (Romanski et al., 1999; for review see Rauschecker and Tian, 2000). The pathway originating from the caudal part of the superior temporal cortex (CL region), shown to be selective for the spatial properties of sounds, projects to the dorsal superior temporal sulcus, posterior parietal areas, and dorsal prefrontal regions, whereas the pathway originating from the anterior areas of the superior temporal gyrus (AL and middle-lateral (ML) regions), projects to the rostral temporal areas, frontal pole, rostral principal sulcus, and ventral prefrontal regions.

Single cells in these cortical areas that receive input from auditory areas respond best to different features of the auditory stimulus. Neurons selectively responsive to vocalizations were found in the ventral prefrontal cortex (Romanski and Goldman-Rakic, 2002). In contrast, neurons responsive to spatial features of auditory stimulation were recorded in the dorsal prefrontal cortex (Azuma and Suzuki, 1984; Vaadia et al., 1986). Discharge of auditory responsive neurons in the temporo-parietal association cortex was dependent on the spatial source of the sound and most of the auditory responses were elicited by natural sounds (Leinonen et al., 1980). The lateral intraparietal area also has been shown to contain neurons with spatially tuned auditory responses (Mazzoni et al., 1996). The responsiveness of auditory neurons in both the prefrontal and parietal cortices is dependent on the behavioral significance of the stimulus; that is, the neurons exhibit stronger responses to active localization or memory tasks than to detection or simple fixation tasks (Vaadia et al., 1986; Grunewald et al., 1999; Linden et al., 1999).

As in the monkey, there is some evidence that the human auditory system also contains functionally dissociable pathways for processing spatial and nonspatial information. Some investigators obtained support for this dissociation in temporal auditory cortex (Baumgart et al., 1999; Belin et al., 2000, 2002; Shah et al., 2001), parietal cortex (Weeks et al., 1999), and prefrontal cortex (Alain et al., 2001), although others have found no clear evidence for such domain specificity in the auditory system (Bushara et al., 1999; Maeder et al., 2001; Zatorre et al., 1999, 2002). The nature of the dissociation as well as the functional neuroanatomy of this possible auditory domain-specificity is therefore not yet clear, and none of these studies provide evidence regarding which cognitive operations required by the tasks were actually responsible for the observed dissociations.

In the present work, we used functional magnetic resonance imaging to study working memory for the location and identity of human voices in an attempt to determine whether the neural system for auditory working memory in humans, like the one for visual working memory (e.g. Courtney et al., 1996, 1998; Sala et al., 2003), exhibits a functional dissociation for spatial and nonspatial information. We used voices because prefrontal neurons in monkeys were shown to respond better to natural sounds or monkey vocalizations than to pure tones (Azuma and Suzuki, 1984; Romanski and Goldman-Rakic, 2002), and also because the anterior part of the superior temporal sulcus (STS)

in humans has been shown to exhibit selective activation for voices, leading to the suggestion that this region may be analogous to the face sensitive area in the fusiform gyrus which is a part of the ventral visual pathway (Belin *et al.*, 2000, 2002; Shah *et al.*, 2001). The subjects performed a delayed recognition task for human voices and voice locations and a sensory-motor control task. To find out whether a functional dissociation between voice and location recognition might occur during specific phases of working memory, we performed separate analyses of task-related activations evoked during the sample, delay, and test periods of the two memory tasks.

Materials and Methods

Subjects

Fourteen right-handed subjects (10 females) between the ages of 18 and 27 years (mean 22 years) participated in the study. The subjects were native English speakers and were screened for mental and physical health. They had no history of head injury, or of drug or alcohol abuse, and no current use of medications that affect central nervous system or cardiovascular function. The subjects gave written informed consent, and were paid 50 USD for participating in the experiment. The experimental protocol was approved by the Review Board on the Use of Human Subjects of the Johns Hopkins University and by the Joint Committee on Clinical Investigations of the Johns Hopkins Medical Institutions.

Stimuli

Voice samples consisted of pairs of words. The first word was a two-syllable adjective, and the second word, a five-syllable noun. The samples were recorded in a sound-proof room using CSL software (Sensimetrics Corporation, Somerville, MA, USA). The sampling rate was 44.1 KHz. The targeted words were situated within a sentence ('John says that [further consideration] is important') to encourage natural speech, and the speakers were instructed to read the sentences in a neutral tone. The pair of targeted words was always situated in the same position within a sentence. The speakers read each sentence twice during the recording. Ten pairs of words were recorded, and three of them ('further consideration', 'simple inauguration', and 'constant unreality'), excerpted from the recorded sentences, were chosen for use in the study. Eight female voices were recorded. The speakers were native English speakers. The mean durations of the three word pairs were 1316 ms (SD 106 ms), 1378 ms (SD 59 ms), and 1333 ms (SD 62 ms), respectively. There were no significant differences in duration between the three pairs [$F(2,21)$ = 1.33, P = 0.29]. The energy levels (db) of voice samples were normalized using CSL/ASPP software.

The voice samples were transformed with head related transfer functions (HRTFs) to create localizable stereo stimuli for presentation through headphones (TDT-PD1 system; Tucker-Davis Technologies). Stimuli were presented at eight possible locations around the head. The coordinates of sounds locations, from the center of the head at nose level, were the following (azimuth/elevation in degrees from straight ahead): 0/40, –30/30, –40/0, –30/-30, 0/-40, 30/-30, 40/0 and 30/30. We created individualized HRTFs for seven different head sizes and measured the dimensions of each subject's head to find the best HRTF for each subject. In the magnet, the stimuli were presented through air conduction headphones.

For the control task, the auditory stimuli were phase-scrambled in the Fourier domain, maintaining frequency information and stimulus amplitude envelopes equal to those in the memory tasks. The phase-scrambled voices were presented simultaneously from four randomly selected locations, and thus were neither identifiable nor localizable.

Each location and voice was presented 24–26 times during the experiment. Before the experiment, the subjects heard each location and voice once to gain familiarity with the stimuli, and once or twice more during the memory task training.

Visual stimuli (e.g. trial instructions and fixation cross) were presented using an LCD projector, located outside of the scanning room, connected to a Power Macintosh G3 computer running SuperLab software. The stimuli were projected on a rear projection screen mounted inside the bore of the magnet, behind the subject's head. Subjects viewed the stimuli through a mirror mounted at the top of the head coil.

Tasks

Two working memory tasks and a control task (Fig. 1) were presented in a delayed recognition paradigm in which subjects were instructed to remember either locations or voices, or neither. One second before each trial, the subjects were presented with an instruction image (for 0.5 s) consisting of the word 'place' (for the location task), 'voice' (for the voice task), or 'none' (for the control task) indicating which task was to be performed. In the location task, the subjects were to memorize the auditory location independent of speaker or words spoken, and in the voice task, the speaker of the sample independent of auditory location or words spoken. The sample was presented for ~1.5 s followed by a memory delay of 4.5 s during which the subjects saw a blank screen with a fixation cross. Then, a test stimulus was presented for ~1.5 s during which time the subject indicated with a left or right button press whether or not the test stimulus was the same as the one in the sample period. Each subject was allowed to choose whether the right or left hand would correspond to the 'match' response. The other hand would be used for the 'no match' response. Responses were made with left or right thumb presses of hand held button boxes that were connected via a fiber optic cable to a Cedrus RB-6 × 0 Response Box. The recorded words presented during the test period never matched the words presented during the sample period. Also, for the voice task, the auditory location presented during the test period never matched the location presented during the sample period. Similarly, for the location task, the voice presented during the test period never matched the voice presented during the sample period. Following each trial there was an intertrial interval of 3.0 s. Subjects also performed a sensorimotor control task with no mnemonic demand. For this task, the scrambled stimuli were presented with the same timing as in the memory tasks but the subjects were instructed that they need not remember the locations or voices, but simply press both buttons when the test stimulus was played.

During the scanning, six runs were conducted. In each run, both memory task conditions were presented in four alternating blocks of four trials each. Each block of four memory task trials was preceded and followed by one control trial. Thus, in each run, there were 8 memory test trials of each information type and eight control trials. The order of tasks was counterbalanced across runs within each subject, and the order of runs was counterbalanced across subjects.
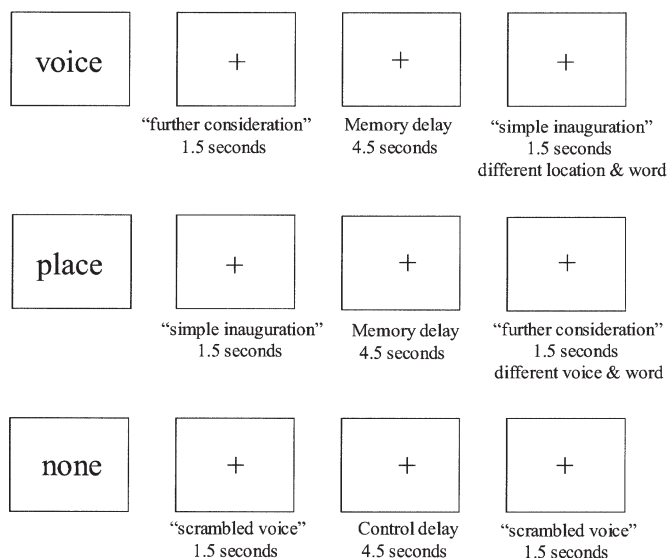
Figure 1. Illustration and timing of the delayed recognition and control tasks.

The reaction times and match/no-match responses were recorded during the scanning. After the scanning, each subject was asked to fill out a questionnaire rating the difficulty of each task and mnemonic strategies used in his/her task performance.

FMR imaging and data analysis

MR-images were acquired with a 1.5 Tesla Philips Gyroscan ACS-NT MR scanner (Philips Medical Systems). A T_1-weighted structural image (70 axial slices, 2.5 mm, no gap, T_R = 20 ms, T_E = 4.6 ms, flip angle = 30°, matrix 256 × 256, FOV = 230 mm) was obtained before the functional scanning. During the performance of the tasks, subjects underwent T_2*-weighted interleaved gradient-echo, echo-planar imaging (21 axial slices, 5 mm thickness, no gap, T_R = 1500 ms, T_E = 40 ms, flip angle = 70°, matrix 64 × 64, FOV = 230 mm). The images were phase-shifted using Fourier transformation to correct for slice acquisition time, then motion-corrected using automatic image registration (AIR) software (Woods et al., 1998), and analyzed separately for each subject using multiple regression (Friston et al., 1995; Ward, 2001) with Analysis of Functional NeuroImages (AFNI) software (Cox, 1996). Changes in neural activity were modeled as square-wave functions matching the time course of events of experimental tasks. These square-waves were convolved with a gamma function model of the hemodynamic response using the following values: 2.0 s for lag, 3.0 s for rise time, and 5.0 s for fall time to create the regressors of interest in the multiple regression analysis. Additional regressors were included to model sources of variance not related to the experimental manipulations (mean intensity between and linear drift within time series). Both memory task conditions (location and voice) were separately contrasted to the control task, and to each other, for each of the three main events of the tasks (sample, delay, and test). Each of these contrasts resulted in a Z-map for each subject.

Z-maps were registered into the Talairach coordinate system (Talairach and Tournoux, 1988) and resampled to 1 mm³. Average Z-maps were computed by dividing the sum of Z-values by the square root of the sample size using AFNI software (Cox, 1996). All tests of voxelwise significance were held to a Z threshold of 2.33, corresponding to a $P < 0.01$, and corrected for multiple comparisons (experiment-wise $P < 0.05$) using a measure of probability that uses the individual voxel Z score threshold and the number of contiguous significant voxels. Based upon a Monte Carlo simulation run via AFNI (Ward, 2000), it was estimated that a 387 mm³ contiguous volume (six voxels, each measuring 3.59 mm × 3.59 mm × 5 mm) for the volume of the entire brain would meet the $P < 0.05$ threshold. For the direct comparison between memory tasks, the analysis was restricted to only those voxels showing significantly greater activity for any of the memory tasks versus control. Within this restricted number of voxels, a 258 mm³ cluster size (four voxels) satisfied a 0.05 experiment-wise probability. Activations were anatomically localized in the averaged maps using T1-weighted images.

Frontal Cortex Region of Interest (ROI) Analysis

Based on anatomical hypotheses derived from previous studies of spatial and nonspatial working memory for visual stimuli (Courtney et al., 1998; Sala et al., 2003), ROIs encompassing the anterior inferior frontal gyrus and anterior insula (IFG/Insula), middle and posterior IFG (IFG), anterior middle frontal gyrus (MFG), and superior frontal sulcus (SFS/PreCG)/precentral gyrus (SFS/PreCG) were drawn in both hemispheres of a Talairach transformed brain according to Brodmann areas (BAs) and anatomical landmarks of the Talairach (Talairach and Tournoux, 1988) and Damasio (1995) brain atlases. The IFG/Insula ROI included BAs 45 and 47 of the IFG (z = –5.0 mm to 16.00 mm). The posterior border of the IFG/Insula ROI was the anterior bank of the sylvian fissure (z = –5.0 mm to 12.0 mm) and the anterior bank of the precentral sulcus (PreCS) (z = 12.0 mm to 16.00 mm). The anterior border of the IFG/Insula ROI was the inferior frontal sulcus (IFS). The IFG ROI included BAs 44 and 45 of the IFG (z = 17.0 mm to 34.0 mm). The posterior border of the IFG ROI was the anterior bank of PreCS, and the anterior border was the posterior bank of IFS. The anterior MFG ROI included BAs 46 and 10 of the MFG (z = 5.0 mm to 23.0 mm). The posterior border of the MFG ROI was the anterior bank of IFS, and the anterior border was the posterior bank of the superior frontal sulcus (SFS). The SFS/PreCG ROI included the SFG within ~6 mm of either

side of the SFS (z = 35.00 mm to 63.00 mm) and BA 6 of the PreCG (z = 44.00 mm to 63.00 mm). The posterior border of the SFS/PreCG ROI was the anterior bank of the central sulcus (CS), and the anterior border was the posterior bank of PreCS.

For each ROI, the number of voxels significantly activated (not corrected for multiple comparisons) in each of the three main periods of the memory tasks relative to the corresponding period in the control task was computed for each subject. The number of significantly activated voxels was then normalized by dividing by the total number of voxels in each ROI. In addition, for each ROI, the signal intensities (β-coefficients) of the significantly activated voxels (corrected for multiple comparisons), determined as described above, were computed for each subject. Analysis of variance for repeated measures with subject as a random factor (BMDP2v, BMDP Statistical Software, Inc., Release 7.1) was used to test the main effects and interactions of task, event, hemisphere, and brain region on both the number of suprathreshold voxels and β-coefficients. A pairwise t-test was then used to test the effect of task on the number of activated voxels and β-coefficients separately for each ROI.

Response Topography Correlation Analysis

In the group average Z maps, clusters of voxels that were activated in any of the planned contrasts – location versus control, voice versus control, or both tasks combined versus control – were assigned to six broadly defined anatomical regions: right and left lateral frontal cortex, right and left posterior parietal cortex, and right and left anterior parietal/temporal cortex. These regions are shown in Figure 7. Within each of these regions, for each subject, voxels were identified that were significantly positively activated for that subject individually, in any of the same contrasts. Within each subject, within each of the three regions (frontal, parietal, and temporal) these voxels were ordered hierarchically first by ventral to dorsal, then by anterior to posterior, then by left to right, to create a 'voxel index'. The beta weights as a function of voxel index for each WM task thus became a single metric for the response topography within each region. The multiple regression was re-run with separate regressors for either odd or even numbered blocks of each task. Correlations were calculated between the response topography on odd blocks and the response topography on even blocks of the same WM task, for a measure of within-task consistency of the topography. Correlations were also calculated between the response topography on odd (even) blocks of one WM task and the response topography on odd (even) blocks of the other WM task, for a measure of between-task consistency of the topography. Correlation coefficients were converted to Z scores and t-tests were performed to test whether or not the response topography within each region was more highly correlated within task than between tasks.

Results

Behavioral Results

The subjects were equally accurate in both memory task conditions. The percentage of correct responses for the location task was 83%, and for the voice task, 84%. The reaction times were significantly faster for the location (1869 ms) than for the voice (1956 ms) task ($P < 0.05$). The subjects evaluated both tasks equally difficult to perform. Subjects reported having used several different memory strategies: Visual, verbal, and auditory imagery to remember the locations, and mainly auditory imagery but also verbal strategies to remember the voices.

fMRI Results

Voxelwise Multiple Regression

Location and Voice Task Activations Relative to Control.

Sample Period (Table 1 and Fig. 2). For the location samples, activation was detected in the left superior temporal sulcus/gyrus (STS/STG) and in the left inferior parietal lobe/

postcentral gyrus (IPL/PostCG). For the voice samples, there was bilateral activation of STS/STG and STG/Insula.

Delay Period (Table 2 and Fig. 3). Several temporal, parietal, and frontal regions were activated during the delay period of the tasks. In the temporal lobe, the right STG and bilateral STS/middle temporal gyrus (MTG) were activated during voice delays. In the parietal lobe, the right IPL and bilateral superior parietal lobe (SPL) were activated only during location delays, whereas the left IPL was activated during both delays. Finally, in the frontal lobe, the anterior middle frontal gyrus (MFG) was activated during location delays, while the inferior frontal gyrus/Insula (IFG/Insula), IFG, superior frontal sulcus/precen-

tral gyrus (SFS/PreCG), and the medial part of the superior frontal gyrus (SFGm) were activated during both delays.

Test Period (Table 3 and Fig. 4). Several temporal, parietal, and frontal regions were also activated during the test period of the tasks. In the temporal lobe, the STS/STG was bilaterally activated by both tasks. In the parietal and frontal cortices, the IPL, SPL, IFG/Insula, IFG, anterior MFG, and SFGm were activated by both tasks.

Direct Voxelwise Comparisons: Location > Voice and Voice > Location (Table 4 and Fig. 5). Direct voxelwise comparisons between the two tasks revealed no significant differences

Table 1
Sample activity for Locations and Voices versus Control

Area	Location > control				Voice > control			
	x, y, z	Peak Z	Mean Z	Spatial extent (mm³)	x, y, z	Peak Z	Mean Z	Spatial extent (mm³)
Temporal								
STS/STG	−55, −22, −2	4.25	2.89	493	60, −22, −1	5.02	3.09	1171
					−56, −11, −2	4.63	2.81	512
STG/insula					46, 11, 0	4.72	2.91	500
Parietal					−50, 4, −1	4.47	2.80	602
IPL/PostCG	−33, −32, 43	4.27	2.71	462				

Areas of significant activity, the peak and mean Z values, the spatial extent of a given activity, and the Talairach coordinates of maximum Z value within each region during the sample period of the memory tasks relative to control task.

Table 2
Delay activity for Locations and Voices versus Control.

Area	Location > control				Voice > control			
	x, y, z	Peak Z	Mean Z	Spatial extent (mm³)	x, y, z	Peak Z	Mean Z	Spatial extent (mm³)
Temporal								
STG					49, −22, 6	5.65	3.00	418
STS/MTG					−55, −51, 3	4.64	2.88	973
					47, −41, 4	4.30	2.85	695
Parietal								
IPL	−41, −55, 44	4.98[a]	2.97	3005	−33, −57,38	5.19	2.95	1042
	31, −50, 36	5.25	3.02	778				
SPL	−7, −71, 45	4.88[a]						
	14, −73, 41	5.34	2.97	942				
Frontal								
IFG/Insula	−40, 21, 15	4.83	3.03	1200	−45, 18, 14	5.59[b]	3.12	10761
	28, 25, 8	4.94	2.91	522	35, 27, 5	4.58	3.00	525
IFG	−40, 11, 34	6.17	3.11	4773	−47, 6, 33	5.66[b]		
	40, 6, 25	6.50	3.11	1968	43, 3, 25	5.58	3.04	2021
Ant. MFG	−38, 46, 15	4.10	2.81	471				
SFS/PreCG	−27, −16, 49	6.49	3.24	2496	−39,−11,46	6.17[b]		
	35, −8, 46	5.41	3.08	2362	30,−11, 46	6.11	3.05	946
SFGm	3, 2, 52	5.59	3.29	3205	−5, 6, 46	6.78	3.46	5487

Areas of significant activity, the peak and mean Z values, the spatial extent of a given activity, and the Talairach coordinates of maximum Z value within each region during the delay period of the memory tasks relative to control task.
[a]The cluster included two separate activation loci with local maxima in both the left IPL and left SPL.
[b]The cluster included three separate activation loci with local maxima in the left IFG/Insula, left IFG, and left SFS/PreCG.

Table 3

Test Activity for Locations and Voices versus Control.

Area	Location > control				Voice > control			
	x, y, z	Peak Z	Mean Z	Spatial extent (mm³)	x, y, z	Peak Z	Mean Z	Spatial extent (mm³)
Temporal								
STS/STG	−55, −22, −2	7.77[a]	3.20	10079	−55, −21, −2	9.51[b]	3.45	26584
	61, −20, 1	7.46[c]	3.24	22965	61, −21, 1	10.21[d]	3.60	36850
Parietal								
IPL	−31, −62, 40	5.89[e]	3.12	4122	−34, −69, 38	7.88	3.22	4384
	26, −75, 40	5.83[f]	3.07	6603	31, −67, 35	5.66	3.06	3977
SPL	−12, −76, 39	5.40[e]			5, −76, 34	5.67	3.10	1241
	−22, −61, 63	6.14	3.33	494				
	14, −71, 37	5.09[f]						
Frontal								
IFG/Insula	−41, 12, 5	5.28[a]			−33, 13, 8	7.83[b]		
	48, 11, 4	5.64[c]			46, 11, 0	10.18[d]		
IFG	−45, 4, 31	6.23	3.03	1783	−44, 8, 30	7.69[b]		
	50, 14, 23	6.41[c]			46, 24, 26	7.83[d]		
Ant. MFG	−42, 39, 8	5.75	3.16	629	−41, 42, 10	5.64	3.11	1269
	−36, 48, 14	4.80	2.89	981				
SFGm	0, 15, 47	6.38	3.22	3088	1, 8, 51	8.82	3.59	6972
Cerebellum	−11, −77, −25	4.62	2.96	765	7, −80, −23	5.56	3.02	1839

Areas of significant activity, the peak and mean Z values, the spatial extent of a given activity, and the Talairach coordinates of maximum Z value within each region during the test period of the memory tasks relative to control task.

[a]The cluster included two separate activation loci with local maxima in both the left STS/STG and left IFG/Insula.
[b]The cluster included three separate activation loci with local maxima in the left STS/STG, left IFG/Insula and left IFG.
[c]The cluster included three separate activation loci with local maxima in the right STS/STG, right IFG/Insula and right IFG.
[d]The cluster included three separate activation loci with local maxima in the right STS/STG, right IFG/Insula and right IFG.
[e]The cluster included two separate activation loci with local maxima in both the left IPL and left SPL.
[f]The cluster included two separate activation loci with local maxima in both the right IPL and right SPL.

during the sample period. During the delay period, the left SFS/PreCG and the right SPL were activated more for the location task than for the voice task, but there was no region exhibiting greater activation for voice than for location delays (when corrected for multiple comparisons). During the test period, however, whereas the right SPL was again activated more for locations than for voices, bilateral IFG/Insula was activated more for voices than for locations.

*ROI Analysis in the Frontal Cortex*
The voxelwise regression analysis suggested a dissociation between dorsal (parietal and SFS/PreCG) and ventral (IFG/Insula) cortical areas for spatial versus nonspatial working memory, but this analysis did not show a convincing double dissociation. Such a result does not prove the absence of a functional dissociation, however, and so the data were further analyzed using two other methods. First, because we had an *a priori* hypothesis regarding specific anatomical criteria for defining regions of interest in the frontal cortex, but not in other areas, we performed an ROI analysis only within the frontal cortex on both the number of activated voxels and the signal intensity (β-coefficients).

Sample Period. The ROI analysis during the sample period of the tasks demonstrated that there was a significant main effect of brain region on the number of significantly activated voxels [$F(3,39) = 3.97$, $P < 0.05$] and on signal intensities

[$F(3,39) = 11.56$, $P < 0.001$] but no main effect of task nor interaction between the task and brain region.

Delay Period. During the delay period, there was a significant main effect of brain region [$F(3,39) = 8.97$, $P < 0.005$] and an interaction between task and brain region [$F(3,39) = 6.20$, $P < 0.005$] on the number of suprathreshold voxels. This number was significantly greater for voice than for location delays in the left IFG/Insula (0.034 versus 0.022 [number of activated voxels divided by the total number of voxels in the ROI], respectively, $P < 0.01$) and the left IFG (0.071 versus 0.053, $P < 0.05$). An interaction between task and brain region during the delay period of the tasks was significant also for signal intensities (β-coefficients) of activated voxels [$F(3,39) = 4.49$, $P < 0.05$]. Signal intensity was significantly greater for voice than for location delays in the left IFG/Insula (0.0056 versus 0.0040, $P < 0.05$). Conversely, signal intensity of activated voxels was significantly greater for location than for voice delays in the right SFS/PreCG (0.0060 versus 0.0033, $P < 0.05$) (Fig. 6).

Test Period. For the test period, there were significant main effects of task [$F(1,13) = 9.05$, $P < 0.05$] and brain region [$F(3,39) = 12.95$, $P < 0.0005$] and an interaction between task and brain region on the number of suprathreshold voxels [$F(3,39) = 9.97$, $P < 0.0005$]. This number was significantly greater for voice than for location in the left (0.059 versus 0.034, $P < 0.005$) and right (0.065 versus 0.043, $P < 0.05$) IFG/Insula and the left (0.089 versus 0.049, $P < 0.001$) and right
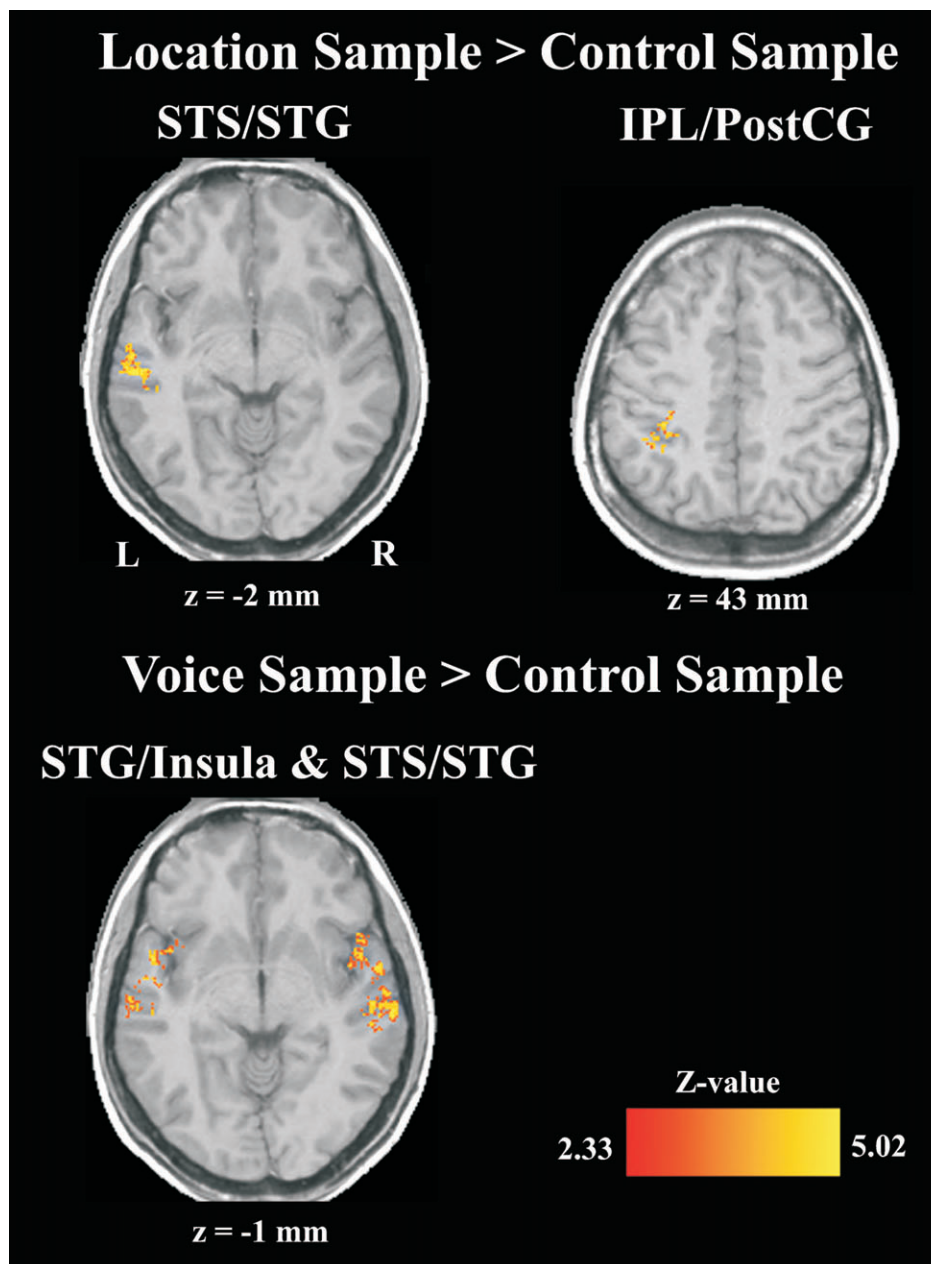
Figure 2. Cross-subject average statistical maps of activation during the sample period of the location and voice tasks (relative to control task) overlayed on a Talairach normalized anatomical image.

(0.102 versus 0.066, $P < 0.001$) IFG. There was also a significant interaction between task and brain region [$F(3,39) = 3.19$, $P < 0.05$] on signal intensities. Signal intensities were significantly greater for voice than for location in the left IFG/Insula (0.0161 versus 0.0099, $P < 0.05$), right IFG/Insula (0.0136 versus 0.0092, $P < 0.05$), and right IFG (0.0155 versus 0.0116, $P < 0.05$).

Comparisons across Sample, Delay, and Test Events. The results obtained from multiple regression and ROI analyses suggest that the nature and magnitude of spatial/nonspatial dissociation may be different at different times during the performance of working memory task. Therefore, we also performed a 4-way ANOVA to test main effects and interactions

of task, event within the task, hemisphere, and brain region. The results showed that there was a significant task × event interaction for number of activated voxels [$F(2,26) = 5.23$, $P < 0.05$], although the interaction for signal intensity was not significant.

Functional Topography Correlation Analysis
To test the robustness of these findings in the frontal cortex and to further test for functional topographies in other brain regions where we did not have such specific, anatomically based hypotheses, we performed a correlation analysis on the pattern of activation magnitude within activated clusters in frontal, parietal, and temporal cortices. The analysis is
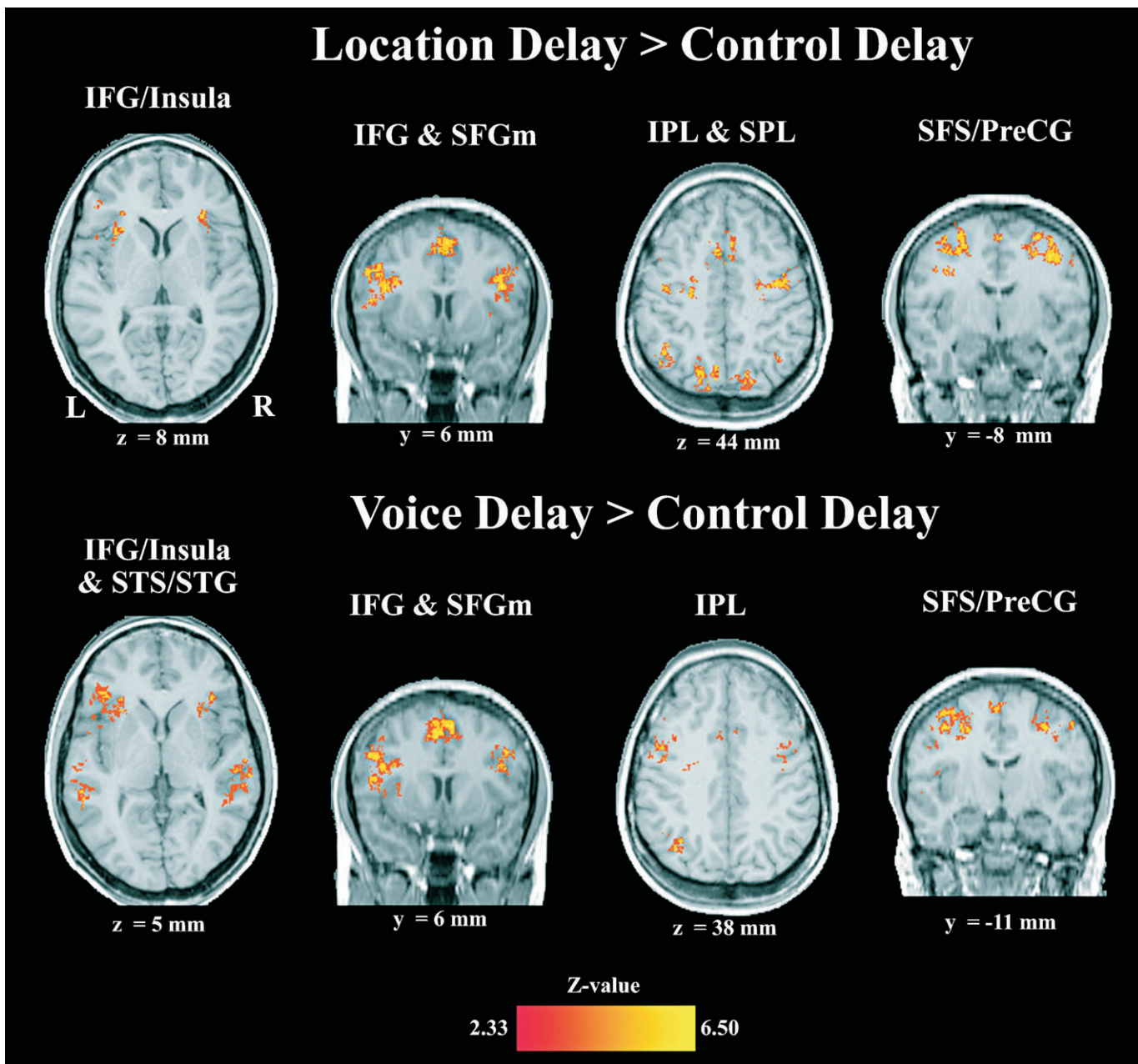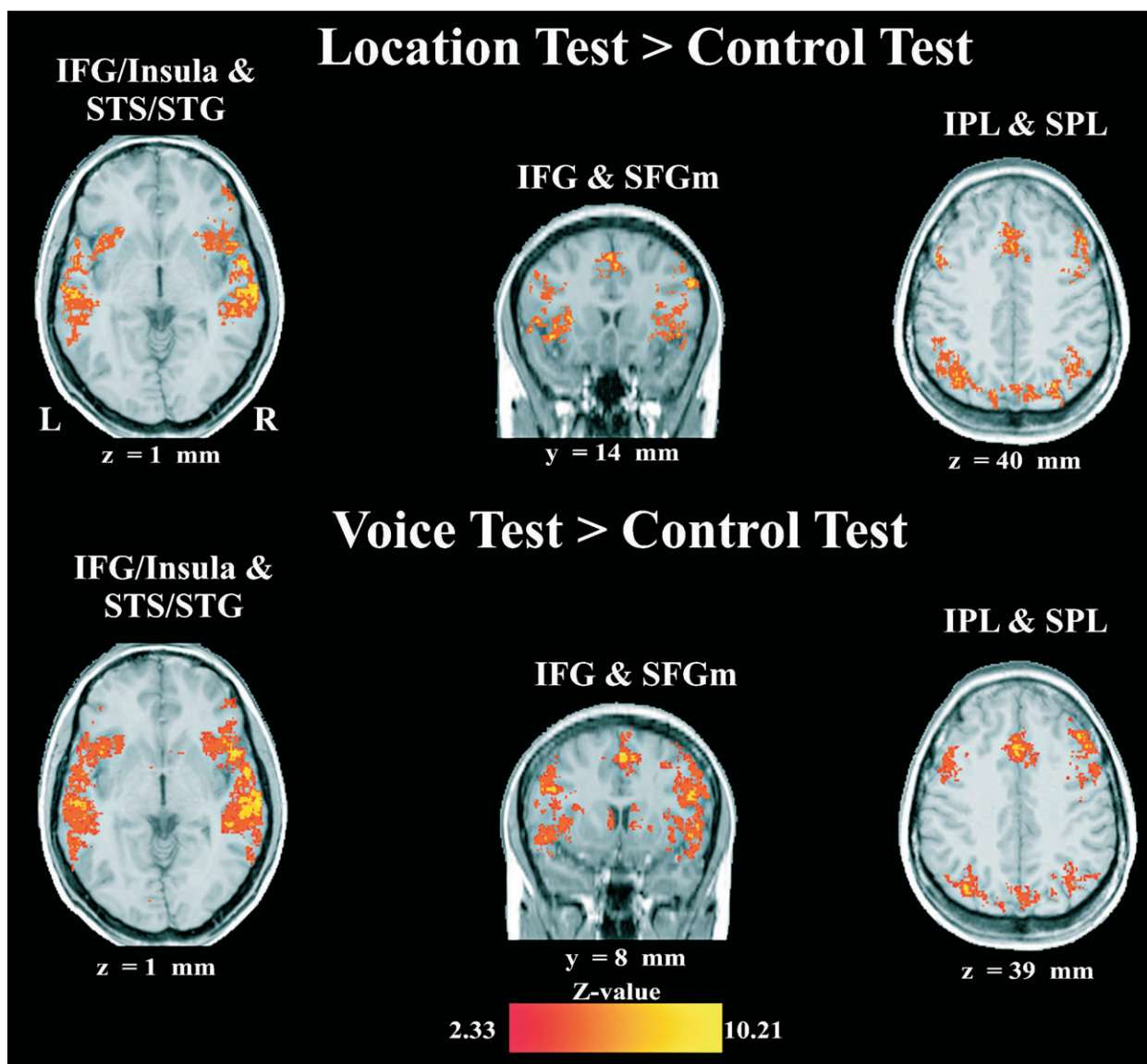
Figure 3. Cross-subject average statistical maps of activation during the delay period of the location and voice tasks (relative to control task) overlayed on a Talairach normalized anatomical image.

described in detail in the Materials and Methods section, and the results for the delay period in two individual subjects are shown in Figure 7.

Sample Period. Functional topographies were not statistically more similar within task (across odd and even blocks) than between task (within odd or even blocks) during the sample period for any of the activated regions, although there was a trend toward this effect in the frontal cortex ($r = 0.36$ versus 0.32, $P = 0.08$). We also calculated the slope of the regression line for the beta coefficients as a function of voxel index for each subject. This is only a rough indicator of the functional topography, because, as can be seen in Figure 7, the plots are highly nonlinear. Nevertheless, for the ventral to dorsal voxel index order, the slopes were significantly different for the

spatial and the identity tasks in the frontal cortex (5.1×10^{-6} and -2.9×10^{-6}, respectively, $P < 0.05$), indicating that the amount of activation for the spatial task increases from ventral to dorsal frontal cortex while the amount of activation for the identity task decreases.

Delay Period. During the delay period, functional topographies were significantly more similar within task than between tasks for left frontal ($r = 0.69$ and 0.56, respectively, $P < 0.005$), right frontal ($r = 0.76$ and 0.68, respectively, $P < 0.05$) and right parietal ($r = 0.74$ and 0.55, respectively, $P < 0.05$) cortices. As illustrated in Figure 7, frontal cortex activation for the location task increased whereas that for the voice identity task decreased with increasing voxel index, indicating a dorsal/ventral spatial/nonspatial functional topography (slopes = 1.86
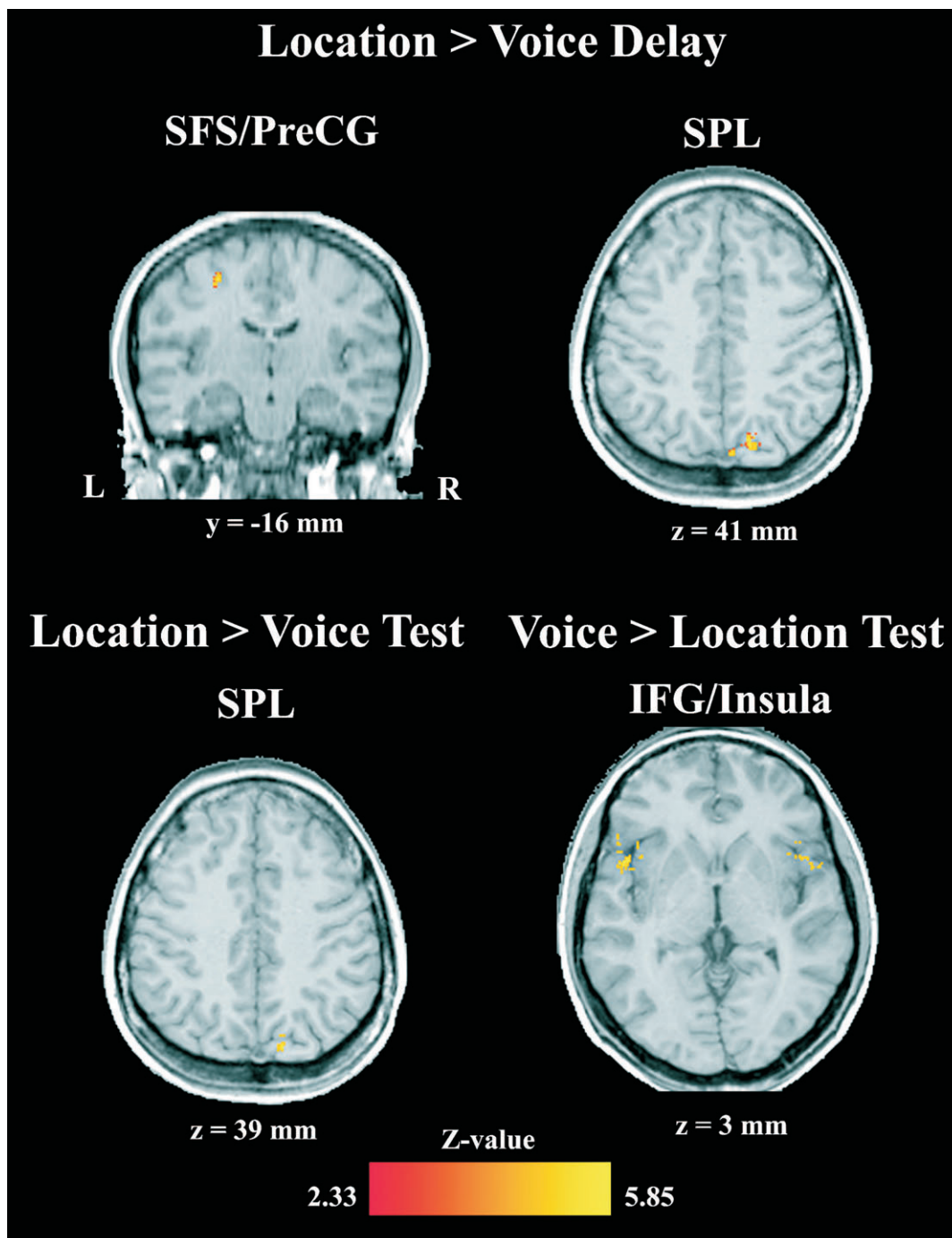
Figure 4. Cross-subject average statistical maps of activation during the test period of the location and voice tasks (relative to control task) overlayed on a Talairach normalized anatomical image.

Table 4
Direct Comparisons between the Memory Tasks

Area	Location > voice				Voice > location			
	x, y, z	Peak Z	Mean Z	Spatial extent (mm³)	x, y, z	Peak Z	Mean Z	Spatial extent (mm³)
Sample								
No statistically significant differences in direct comparisons								
Delay								
SFS/PreCG	−26, −16, 50	4.60	2.96	380				
SPL	14, −73, 41	5.85	3.50	646				
Test								
SPL	13, −78, 39	4.46	3.02	404				
IFG/insula					−46, 10, 3	4.48	3.01	716
					42, 12, −1	4.33	2.82	501

Areas of significant activity, the peak and mean Z values, the spatial extent of a given activity, and the Talairach coordinates of maximum Z value within each region during the all periods of the memory tasks for locations versus voices.

Figure 5. Cross-subject average statistical maps of direct comparisons between activations during the delay and test periods of the location and voice tasks overlayed on a Talairach normalized anatomical image. There were no significant differences during the sample period.

$\times 10^{-6}$ versus -1.62×10^{-6} respectively, $P < 0.05$). Although the spatial task tended to produce greater activation than did the voice identity task across all activated portions of parietal cortex, the dissociation between the tasks was greatest in the most superior portion of this region (slopes = 1.30×10^{-4} versus 4.12×10^{-5} respectively, $P < 0.05$). The results of the correlation analysis are independent of the particular ordering chosen to define the voxel index. If the voxels are ordered first from posterior to anterior instead of from ventral to dorsal, and similar plots are prepared for the activations in the temporal

region of the same subjects illustrated in Figure 7, there appears to be greater activation for the identity task in the anterior portion of the temporal region, consistent with previously reported results (for review see Rauschecker and Tian, 2000). However, neither left nor right temporal cortex showed a consistent functional topography across subjects with this analysis (r\{within/between\} = \{0.70/0.72\}, $P = 0.3$ and $r = $ \{0.76, 0.73\}, $P = 0.3$, respectively.

Test Period. During the test period, functional topographies were not statistically more similar within task than
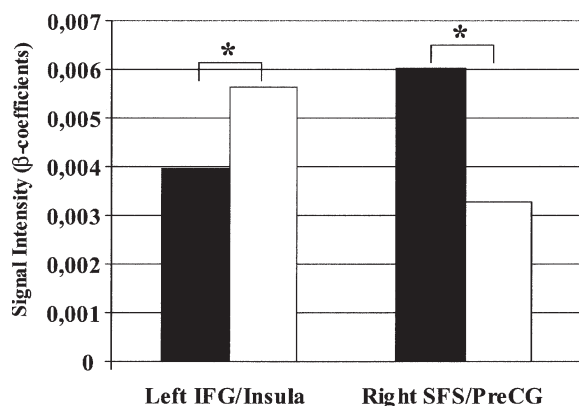
Figure 6. Results of the ROI analysis showing a double dissociation in signal intensity during the delay period of location (black bars) and voice (white bars) tasks in the left IFG/Insula and the right SFS/PreCG, $P < 0.05$.

between task for any of the activated regions. However, the slopes were significantly different for the spatial and the identity tasks in the right frontal cortex (3.5×10^{-6} and -9.6×10^{-6} respectively, $P < 0.05$), indicating again that the amount of activation for the spatial task increases from ventral to dorsal frontal cortex while the amount of activation for the identity task decreases.

Discussion

The present results show that working memory maintenance for voices and auditory locations activates a distributed neural network including temporal, parietal, and frontal regions. Taken together, the results from the three different types of statistical analyses indicate that the magnitude of activation within these activated areas shows a different functional topography depending on the type of information being maintained. Activation in the dorsal frontal cortex (SFS/PreCG) and posterior parietal cortex (SPL) was greater for location delays than for voice delays. Conversely, ventral frontal regions (IFG/Insula and lFG) were more active for voice than for location delays. The present findings, together with previous research, indicate that, during auditory working memory, maintenance of spatial or nonspatial information modulates activity in dorsal and ventral frontal cortex, respectively. These results support the idea that the frontal cortex is organized, in part, according to the type of information being maintained in working memory (Wilson *et al.*, 1993; Levy and Goldman-Rakic, 2000).

Previous neuroimaging studies on spatial and nonspatial auditory processing have not provided evidence regarding which cognitive operations required by the tasks were responsible for the observed spatial/nonspatial dissociations (e.g. Weeks *et al.*, 1999; Alain *et al.*, 2001; Maeder *et al.*, 2001). The current study suggests that the magnitude of the dissociation is greatest during maintenance in working memory (i.e. delay period), less during recognition or retrieval (test period), and least during encoding (sample period). The reason for this result is not entirely clear. Examination of the data suggests that the variance in the beta coefficient estimates was greater for the sample and test periods than during the delay, possibly because of intersubject variability in the hemodynamic lag. It also appears that the spatial extent of the proposed spatial/nonspatial functional topography may be smaller in parietal and temporal areas than in prefrontal cortex. Therefore, inter-

subject anatomical variability would interfere more with our ability to detect such a functional topography in the former areas. In addition, it may be that during stimulus presentation, both spatial and nonspatial information are processed, but because only the task-relevant information is actively maintained during the delay, the difference between the tasks becomes more pronounced during this time. This would also help explain why the dissociation was most robust in frontal cortex rather than in posterior areas. Posterior areas would be expected to show a greater dissociation if the differences in activation pattern reflected attentional modulation during stimulus presentation rather than working memory maintenance.

Unlike previous studies of spatial versus nonspatial auditory working memory, the present study used the identity of human voices as the nonspatial information to be remembered. Vocalization and natural sounds have been shown to elicit strong neuronal responses throughout the auditory system, including the temporal, parietal, and frontal cortices (Leinonen *et al.*, 1980; Azuma and Suzuki, 1984; Romanski and Goldman-Rakic, 2002). As with human faces, a human voice contains information about the identity of a person and, thus, it can be considered as an 'auditory face' (Belin *et al.*, 2000, 2002). In monkeys, it has been shown that neurons sensitive to monkey vocalization were located in the ventral prefrontal cortex (Romanski and Goldman-Rakic, 2002). In the visual system, a clear ventral/dorsal dissociation in the prefrontal cortex was demonstrated earlier using faces and locations of faces as memoranda (Courtney *et al.*, 1996, 1998; Sala *et al.*, 2003). Working memory maintenance of face identity preferentially activated the inferior and middle frontal gyri, whereas maintenance of face locations preferentially activated the superior frontal sulcus (Courtney *et al.*, 1996, 1998; Sala *et al.*, 2003). This dorsal/ventral dissociation for visual locations and objects may be greater for faces than for other objects, but it is not specific to faces, as other objects show the same dissociation (Sala *et al.*, 2003). Therefore, it is reasonable to presume that the dissociation observed in the current study is a general spatial versus nonspatial distinction and is not specific to voices. Indeed, the same dissociation was observed by Alain *et al.* (2001) and Arnott *et al.* (2002) using synthesized noise bursts. Ventral prefrontal regions have also been shown to be recruited by other types of nonspatial auditory tasks such as melodic, phonemic and pitch discrimination (Zatorre *et al.*, 1992, 1994; Hsieh *et al.*, 2001).

Previous research regarding spatial and nonspatial auditory perception, attention, and working memory has yielded seemingly contradictory results regarding whether there are dissociable neural systems for the different information domains. In one study, the right auditory cortex was shown to exhibit greater activity for moving than for stationary sounds (Baumgart *et al.*, 1999). However, other studies have not found differential activity in the auditory cortex during active localization of sounds relative to passive listening (Bushara *et al.*, 1999; Weeks *et al.*, 1999). Recently, it was shown that the posterior superior temporal gyrus (STG) was activated by simultaneously presented spatially and spectrotemporally variable sounds but not by sequentially presented sounds, suggesting that the posterior STG is sensitive to both spatial and spectrotemporal features of sounds (Zatorre *et al.*, 2002). In the present study, although there were no significant differences between the location and voice tasks during the sample period in the voxel-wise multiple regression analysis, inspection of the patterns of
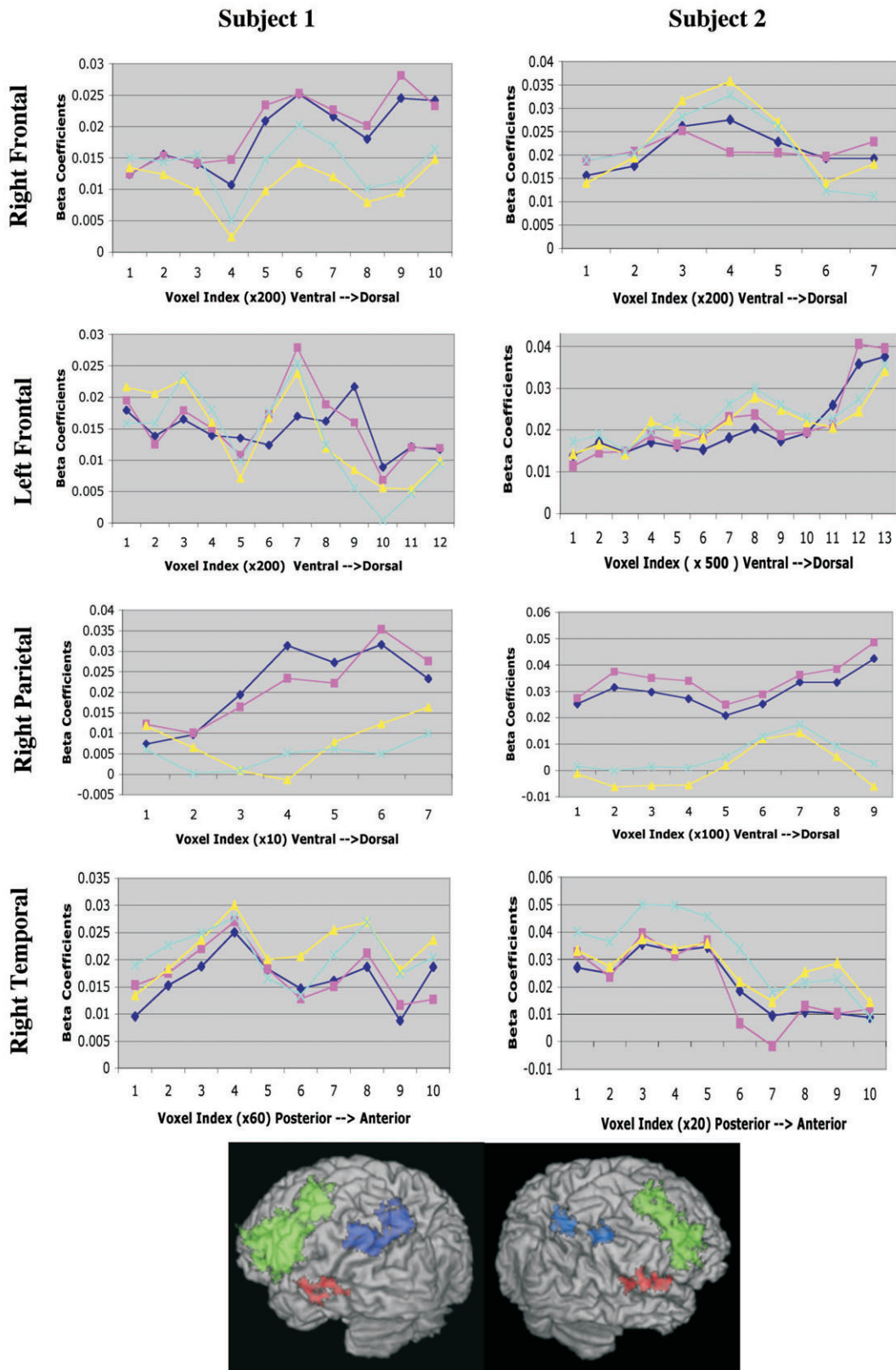
Figure 7. Results of the functional topography correlation analysis for two individual subjects. The brain image shows, in lateral surface projection, the six regions (defined from the group multiple regression analysis) within which the response topographies were analyzed: left and right frontal (green), left and right temporal (red), and left and right parietal (blue). The topographies are plotted with response magnitude (beta coefficient) as a function of voxel index. Pink and dark blue lines illustrate odd and even trials for location task, and yellow and light blue lines odd and even trials for voice task. For the frontal and parietal regions, the voxels are ordered first from ventral to dorsal. For the temporal region, the voxels are ordered first from posterior to anterior. For illustrative purposes, each data point shown is the average of the beta coefficients from 20–500 voxels depending on the total number of significantly activated voxels for that subject within each region.

activation for each memory task versus control suggests that perhaps the voice activation extends further anteriorly than the location activation. Results from individual subjects in the correlation analysis (Fig. 7) also suggest that anterior temporal cortex responds more during the voice identity task than the location task. Such a pattern would be consistent with the organization of auditory cortex that has been found in monkeys, with the AL and ML regions more selective for nonspatial auditory features and the CL region being more selective for auditory locations (for review see Rauschecker and Tian, 2000). Similarly, although the differences were not significant in direct comparisons between the location and voice tasks, the inferior parietal cortex was activated by location samples relative to control samples, but not by voice relative to control samples, which is in line with previous studies showing that the parietal cortex is involved in discrimination and memorizing of audiospatial information (e.g. Bushara *et al.*, 1999; Weeks *et al.*, 1999; Martinkauppi *et al.*, 2000; Zatorre *et al.*, 2002). Therefore, although the current results do not provide direct evidence for a spatial/nonspatial organization in auditory association areas during encoding, they are not inconsistent with this idea.

Only a few neuroimaging studies have compared spatial and nonspatial auditory processing directly (e.g. Weeks *et al.*, 1999; Zatorre *et al.*, 1999; Alain *et al.*, 2001; Maeder *et al.*, 2001). In one study, auditory attention to locations in space and sound frequencies were shown to activate similar cortical regions in temporal, parietal, and frontal regions (Zatorre *et al.*, 1999). On the other hand, three other studies showed anatomically dissociable patterns of activation during sound identification and localization tasks (Weeks *et al.*, 1999; Alain *et al.*, 2001; Maeder *et al.*, 2001). In the study by Weeks *et al.* (1999), the subjects were performing frequency and location discrimination tasks, and the right inferior parietal cortex was shown to be predominantly activated by localization, whereas the left inferior parietal cortex by identification of sounds. Although the primary dissociation in the current study was between right dorsal frontal for the spatial task and left ventral frontal for the nonspatial task, there were no hemispheric laterality differences within parietal cortex, or within ventral or dorsal prefrontal cortex. Alain *et al.* (2001) asked their subjects to perform a delayed comparison task with 1 second delay for locations and frequencies of synthesized sounds. The results showed that the right inferior frontal gyrus was activated more by the pitch than by the location task, whereas the right superior frontal sulcus was activated more by the location than by the pitch task. In the study by Maeder *et al.* (2001), the subjects were also asked to perform a delayed comparison task for locations of noise bursts. In their nonspatial task, the subjects were asked to detect certain environmental sounds (animal cries) among the others (e.g. street, beach, railway station). This study did not reveal as clear a ventral/dorsal dissociation in the frontal cortex as did the study by Alain *et al.* (2001). There were slight differences in locations of peak activities for direct comparisons between the tasks, but ventral and dorsal prefrontal regions were activated for both comparisons. The results of the current study are more similar to those of Alain *et al.* (2001).

The overall dorsal/ventral, spatial/nonspatial functional topography of the frontal cortex appears to be highly similar for auditory and visual working memory (e.g. Levy and Goldman-Rakic, 2000; Sala *et al.*, 2003). Evidence from the monkey suggests that there is an auditory processing domain, separate from the visual processing domain, in the ventral prefrontal cortex. Auditory neurons were located more anteriorly and laterally than were visually responsive neurons (Romanski and Goldman-Rakic, 2002). In humans, however, within the spatial resolution of fMRI, working memory maintenance of faces and of voices appear to activate the ventral frontal cortex similarly (Rämä et al. 2001). Further research is needed to ascertain whether there are two distinct systems for maintenance of visual and auditory information in frontal cortex, both of which show a dorsal/ventral, spatial/nonspatial functional topography, or whether there is a single system for information maintenance independent of stimulus modality.

Notes

Address correspondence to Pia Rämä, Cognitive Brain Research Unit, Department of Psychology, PO Box 9, 00014 University of Helsinki, Finland. Email: prama@cc.helsinki.fi.

References

Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001) 'What' and 'where' in the human auditory system. Proc Natl Acad Sci USA 98: 12301–12306.

Arnott SR, Alain C, Hevenor S, Graham S, Dade LA, Grady C (2002) What, where, and how in the human prefrontal cortex. Program No. 181.1. 2002 Abstract Viewer/Itinerary Planner. Washington, DC: Society for Neuroscience, 2002. Online.

Azuma M, Suzuki H (1984) Properties and distribution of auditory neurons in the dorsolateral prefrontal cortex of the alert monkey. Brain Res 298: 343–346.

Baumgart F, Gaschler-Markefski B, Woldorff MG, Heinze HJ, Scheich H (1999) A movement-sensitive area in auditory cortex. Nature 400: 724–726.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403: 309–312.

Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds.Brain Res Cogn Brain Res 13: 17–26.

Bushara KO, Weeks RA, Ishii K, Catalan M., Tian B, Rauschecker JP, Hallett M (1999) Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. Nat Neurosci 8: 759–766.

Courtney SM, Ungerleider LG, Keil K, Haxby JV (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cereb Cortex* 6: 39–49.

Courtney SM, Petit L, Maisog JM, Ungerleider LG, Haxby JV (1998). An area specialized for spatial working memory in human frontal cortex. Science 279: 1347–1351.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29: 162–173.

Damasio H (1995) Human brain anatomy in computerized images. New York: Oxford University Press.

Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams CR, Frackowiak RSJ (1995) Analysis of fMRI time-series revisited. Neuroimage 2: 45–53.

Grunewald A, Linden JF, Andersen RA (1999) Responses to auditory stimuli in macaque lateral intraparietal area. I. Effects of training. J Neurophysiol 82:330–342.

Hsieh L, Gandour J, Wong D, Hutchins GD (2001) Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. Brain Lang 76: 227–252.

Leinonen L, Hyvärinen J, Sovijärvi ARA (1980) Functional properties of neurons in the temporo-parietal association cortex of awake monkey. Exp Brain Res 39: 203–215.

Levy R, Goldman-Rakic PS (2000) Segregation of working memory functions within the dorsolateral prefrontal cortex. Exp Brain Res 133: 23–32.

Linden JF, Grunewald A, Andersen RA (1999) Responses to auditory stimuli in macaque lateral intraparietal area. II. Behavioral modulation. J Neurophysiol 82:343–358.

Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Pittet A, Clarke S (2001) Distinct pathways involved in sound recognition and localization: a human fMRI study. Neuroimage 14:802–816.

Martinkauppi S, Rämä P, Korvenoja A, Aronen H., Carlson S (2000) Working memory of auditory localization. Cereb Cortex 10: 889–898.

Mazzoni P, Bracewell RM, Barash S, Andersen RA (1996) Spatially tuned auditory responses in area LIP of macaques performing memory saccades to acoustic targets. J Neurophysiol 75: 1233–1241.

Rämä P, Falconero L, Courtney SM (2001) Working memory for faces and voices. Soc Neurosci Abstr 25:81.5.

Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. Proc Natl Acad Sci USA 97:11800–11806.

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 12: 1131–1136

Romanski LM, Goldman-Rakic PS (2002) An auditory domain in primate prefrontal cortex. Nat Neurosci 5: 15–16.

Sala JB, Rämä P, Courtney SM (2003) Functional topography of a distributed neural system for spatial and nonspatial information maintenance in working memory. Neuropsychologia 41: 341–356.

Shah NJ, Marshall JC, Zafiris O, Schwab A, Zilles K, Markowitsch HJ, Fink GR (2001) The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. Brain 124: 804–815.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. Science 292: 290–293.

Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: Analysis of visual behavior (Ingle DJ, Goodale MA, Mansfield RJW, eds). Cambridge: MIT Press.

Ungerleider LG, Haxby JV (1994) 'What' and 'where' in the human brain. Curr Opin Neurobiol 4:157–165.

Vaadia E, Benson DA, Hienz RD, Goldstein MH Jr (1986) Unit study of monkey frontal cortex: active localization of auditory and of visual stimuli. J Neurophysiol 56:934–952.

Ward BD (2000) Simultaneous inference for fMRI data. http://afni.nimh.nih.gov/afni/docpdf/AlphaSim.pdf

Ward BD (2001) Deconvolution analysis of fMRI time series data. http://afni.nimh.nih.gov/afni/docpdf/3dDeconvolve.pdf

Weeks RA, Aziz-Sultan A, Bushara KO, Tian B, Wessinger CM, Dang N, Rauschecker JP, Hallett M (1999) A PET study of human auditory processing. Neurosci Lett 12: 155–158.

Wilson FA, Scalaidhe SP, Goldman-Rakic PS (1993) Dissociation of object and spatial processing domains in primate prefrontal cortex. Science 260(5116):1955–1958.

Woods RP, Grafton S., Holmes CJ, Cherry SR, Mazziotta JC (1998) Automated image registration: I., General methods and intrasubject, intramodality validation. J Comput Assist Tomogr 22: 139–152.

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

Zatorre RJ, Evans AC, Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. J Neurosci 14: 1908–1919.

Zatorre RJ, Mondor TA, Evans AC (1999) Auditory attention to space and frequency activates similar cerebral systems. Neuroimage 10:544–554.

Zatorre RJ, Bouffard M, Ahad P, Belin P (2002) Where is 'where' in the human auditory cortex? Nat Neurosci 5: 905–909.