Behavioral/Systems/Cognitive

Functionally Segregated Neural Substrates for Arbitrary Audiovisual Paired-Association Learning

Hiroki C. Tanabe,[1,2] **Manabu Honda,**[1,3] **and Norihiro Sadato**[1,2]

[1]Division of Cerebral Integration, Department of Cerebral Research, National Institute for Physiological Sciences, Okazaki, Aichi 444-8585, Japan, [2]Research Institute of Science and Technology for Society, Japan Science and Technology Agency (JST), Tokyo 105-6218, Japan, and [3]Precursory Research for Embryonic Science and Technology, JST, Kawaguchi 332-0012, Japan

To clarify the neural substrates and their dynamics during crossmodal association learning, we conducted functional magnetic resonance imaging (MRI) during audiovisual paired-association learning of delayed matching-to-sample tasks. Thirty subjects were involved in the study; 15 performed an audiovisual paired-association learning task, and the remainder completed a control visuo-visual task. Each trial consisted of the successive presentation of a pair of stimuli. Subjects were asked to identify predefined audiovisual or visuo-visual pairs by trial and error. Feedback for each trial was given regardless of whether the response was correct or incorrect. During the delay period, several areas showed an increase in the MRI signal as learning proceeded: crossmodal activity increased in unimodal areas corresponding to visual or auditory areas, and polymodal responses increased in the occipitotemporal junction and parahippocampal gyrus. This pattern was not observed in the visuo-visual intramodal paired-association learning task, suggesting that crossmodal associations might be formed by binding unimodal sensory areas via polymodal regions. In both the audiovisual and visuo-visual tasks, the MRI signal in the superior temporal sulcus (STS) in response to the second stimulus and feedback peaked during the early phase of learning and then decreased, indicating that the STS might be key to the creation of paired associations, regardless of stimulus type. In contrast to the activity changes in the regions discussed above, there was constant activity in the frontoparietal circuit during the delay period in both tasks, implying that the neural substrates for the formation and storage of paired associates are distinct from working memory circuits.

Key words: crossmodal; audiovisual; paired association; learning; functional MRI; regression analysis

Introduction

Learning and memory involve the formation of arbitrary links between information. To explore the neuronal dynamics of active memory (Fuster et al., 2000), the delayed paired-association learning task can be used to investigate the neural networks "related" to the generation, maintenance, recall, and representation of a specific memory (Miyashita, 2000; Miyashita and Hayashi, 2000). The first reported neural correlates of associative long-term memories of randomly assigned visual stimuli pairs were in the monkey inferotemporal (IT) cortex (Miyashita, 1988; Sakai and Miyashita, 1991). During the delay period after the presentation of either member of a paired associate in well trained monkeys, IT cortex neurons were active; this activation pattern represents the memory. Gibson and Maunsell (1997) found such a representation in IT cortex for audiovisual crossmodal associations. Prefrontal cortex neurons are part of the network representing crossmodal associations (Fuster et al., 2000). Using visuo-tactile delayed matching-to-sample tasks, Zhou and Fuster (1996, 1997, 2000) demonstrated sustained activity in neurons in the primary somatosensory area during the delay period after presentation of the visual stimuli associated with the tactile sensation. These findings suggest that the crossmodal association memory is represented by the coactivation of the multiple cortical areas involved in each sensory modality. However, the distribution of the neural substrates of crossmodal association memory and their dynamics during learning remain unclear (Gibson and Maunsell, 1997). Furthermore, whether these neural substrates are specific to crossmodal association memory is yet to be determined.

In the present study, we used functional magnetic resonance imaging (MRI) to investigate the neural substrates of crossmodal paired-association learning in humans by investigating the temporal changes in neuronal activity during learning. We used an audiovisual crossmodal paired-association learning task with prelearned stimuli (see Fig. 1). Different participants completed a control visuo-visual intramodal paired-association learning task. The subjects learned the paired relationship by trial and error. Two stimuli (S1 and S2) with a delay period were presented successively, and subjects responded in a forced-choice manner, followed by feedback (F). Our first prediction was that the expected outcome (i.e., the paired associate of S1) would be represented by the crossmodal activation during the delay period, with activation increasing as learning progressed. Second, we anticipated

Received Feb. 16, 2005; revised April 23, 2005; accepted May 18, 2005.

This work was supported by Grants-in-Aid B#14380380 and S#1710003 for Scientific Research from the Japan Society for the Promotion of Science (to N.S.) and by Special Coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. We are grateful to T. Okada for the task design and the determination of the scan parameters.

Correspondence should be addressed to Norihiro Sadato, 38, Nishigo-naka, Myodaiji, Okazaki, Aichi 444-8585, Japan. E-mail: sadato@nips.ac.jp.

DOI:10.1523/JNEUROSCI.0636-05.2005

that areas showing decreasing F-related activation with learning would represent the neural substrates involved in building the representation. This is because, given feedback, the stimuli linkages should be learned by using the prediction error to update expectations about the outcome (S2) so that the expected outcome converges toward the actual outcome across trials (Rescorla and Wagner, 1972; Pearce and Hall, 1980). The prediction error detected when the F is presented should decrease with learning. Third, we predicted that information about S1 and/or the expected outcome would be temporarily held in working memory; this should be represented by sustained activity during the delay period (Baddeley, 1992; Smith and Jonides, 1998; Hartley and Speer, 2000). This sustained activity should remain unchanged throughout learning. Although our first expectation is specific to audiovisual tasks, the latter two are common to both the audiovisual and visuo-visual tasks.

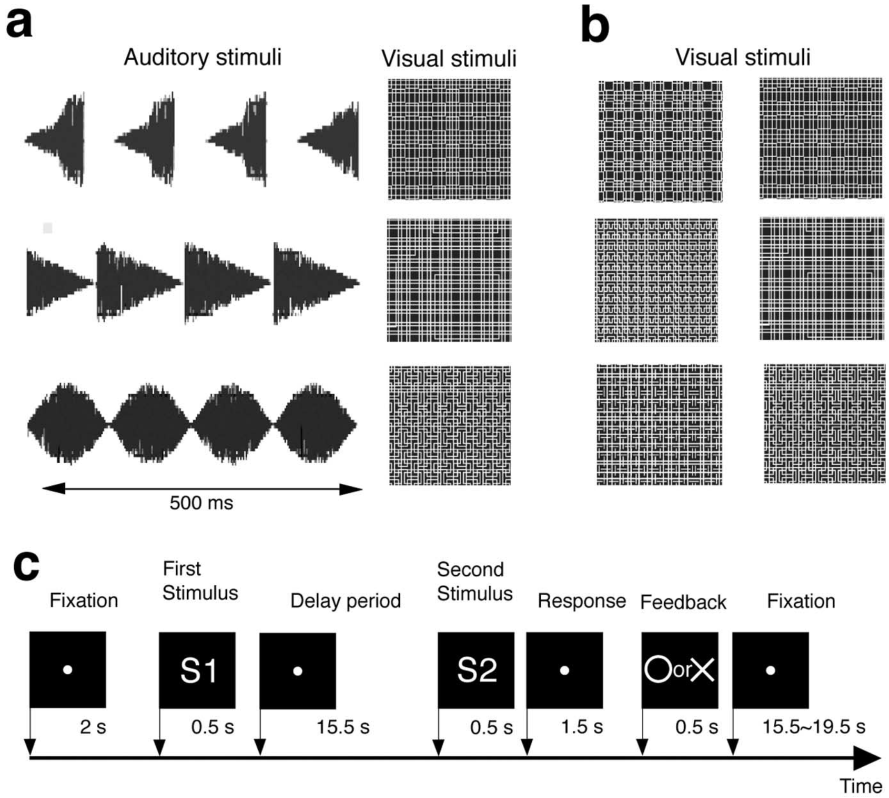

Figure 1. ***a***, Auditory and visual stimuli for the audiovisual task. Auditory stimuli are shown in wave form. ***b***, Visual stimuli for the visuo-visual task. One-half of the stimuli were the same as those for the audiovisual task. These stimuli were divided into two groups (left, VVa; right, VVb). ***c***, Schematic diagram of a paired-association trial. Subjects initially saw one of six stimuli and kept it and/or its associates in mind during a 16 s delay period with a blue fixation stimulus. Then, a second stimulus was shown, which, in the audiovisual task, was of a different modality from the first stimulus. The subject responded with a button press using the right index or middle finger when the fixation stimulus turned red. Visual feedback was presented, enabling subjects to learn the pairs via trial and error. Feedback was presented in the initial six sessions (108 trials). No feedback was shown in the last three sessions.

Materials and Methods

Subjects. Eighteen subjects participated in the audiovisual paired-association learning study (seven females and 11 males; mean age, 26; age range, 21–35). Three subjects were excluded because of poor performance (one female subject) or excessive head motion during the functional MRI scanning (two male subjects), and the data from the remaining 15 subjects were used for additional analysis. In the visuo-visual association learning study, 16 subjects participated, but one female subject was excluded because of poor performance. Therefore, the data from 15 subjects were used for additional analysis. All subjects had normal or corrected-to-normal visual acuity and were right-handed, according to the Edinburgh handedness inventory (Oldfield, 1971). The protocol was approved by the ethical committee of the National Institute for Physiological Sciences, Japan. All subjects gave written informed consent.

Experimental design and task procedure. Subjects completed either the audiovisual or visuo-visual paired-association learning task. Both tasks followed the same procedure but differed in the stimuli used (Fig. 1*a*,*b*). The subjects were asked to identify, by trial and error, three predefined audiovisual pairs (Fig. 1*a*) of nine possible pairs (three auditory and three visual stimuli) in the audiovisual task or visuo-visual pairs (Fig. 1*b*) in the visuo-visual task. The sound stimuli were generated by temporally modulating 500-ms-duration white noise (sampling rate, 44.1 kHz; stereo sound) using Matlab 6.5 (MathWorks, Natick, MA), Sound Builder 3.0 (shareware, developed by Hidaka Ken-ichiro, Japan; http://www.venus.dti.ne.jp/~khidaka/home_en.html) and GoldWave 4.26 (GoldWave, St. John's, Newfoundland, Canada). The sound waves are shown in Figure 1*a* (left). To provide visual stimuli, two-dimensional amorphous texture patterns were downloaded free of charge from http://page.freett.com/amorphis, and their sizes and contrasts were modified. The visual stimuli were 4 × 4° in size and subtended a visual angle of 19 × 14° (Fig. 1*a*, right). Figure 1*b* shows the visual stimuli for the visuo-visual task. One-half of the visual stimuli were the same as those for the audiovisual task (VVa group), and the other stimuli (VVb group) were different from the VVa stimuli.

Presentation 0.50 (Neurobehavioral Systems, Albany, CA) was implemented on a personal computer (Dimension 8200; Dell Computer, Round Rock, TX) for stimulus presentation and response collection. A liquid crystal display projector (DLA-M200L; Victor, Yokohama, Japan), located outside and behind the scanner, projected stimuli through another waveguide to a translucent screen, which the subjects viewed via a mirror attached to the head coil of the MRI scanner. The auditory stimuli were presented via MRI-compatible headphones (Hitachi, Yokohama, Japan). The volume of the sound was adjusted for each subject to an appropriate level for task execution, taking into account the MR scanner noise. Responses were collected via an optical button box (Current Design, Philadelphia, PA).

The task was explained to the subjects in detail, and the subjects recognized all of the auditory and visual stimuli before the scanning session. During the sessions, the subjects were required to direct their eyes toward a fixation point. Each trial consisted of the successive presentation of a "pair" of stimuli (S1 and S2) with a fixed S1–S2 interval (16 s); the duration of each stimulus was 500 ms (Fig. 1*c*). In the auditory–visual (AV) condition, S1 was the auditory stimulus, and S2 was the visual stimulus; these roles were reversed in the visual–auditory (VA) condition. The S2 stimulus subsequently disappeared, and the fixation point turned red, cuing the subject to respond by pressing a preassigned button with the right index or middle finger to report whether the two stimuli (S1 and S2) were a pair or "not a pair." Subjects were asked to perform as quickly and as accurately as possible. Pictorial positive and negative feedback was given in the first 108 trials (six sessions; "learning phase"), 1500 ms after the disappearance of S2. The subjects were asked to correctly pair the stimuli using this feedback information. They were instructed to not use verbalization or labeling strategies to memorize the relationship between the stimuli. No other specific instruction regarding the delay pe-

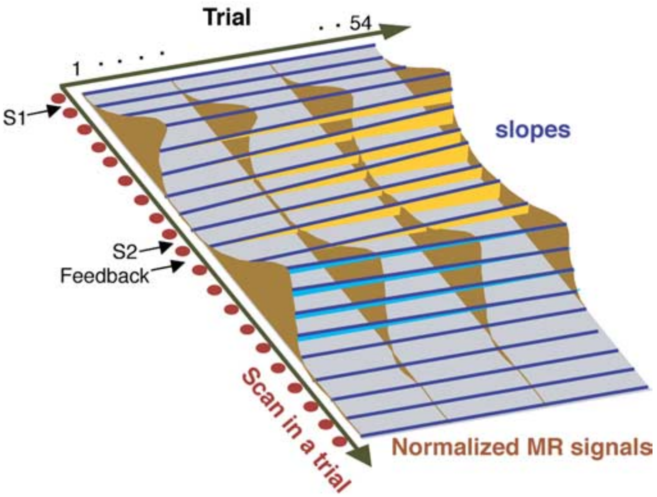

Figure 2. Schematic of the linear regression analysis. Red dots denote a scan in a trial (1–20 scans). Slopes were calculated independently for every 20 scan points. Orange indicates a positive slope (signal value increases throughout the trials), whereas sky blue indicates a negative slope (signal value decreases).

riod was given. For the final 54 trials (three sessions; "learned phase"), no feedback was presented, because the subjects were expected to have identified all three of the audiovisual pairs. The intertrial intervals were pseudorandomized and were 15.5, 17.5, or 19.5 s in length; 10 s were added every six trials to give an obvious baseline measure. The same procedure was performed in the visuo-visual task. The visual S1 was selected from the VVa group (Fig. 1*b*, left), and the visual S2 was chosen from VVb group (Fig. 1*b*, right) for the visual–visual 1 (VV1) condition and vice versa for the VV2 condition. Hence, the VV1 corresponded to the AV condition, and the VV2 corresponded to the VA condition.

A total of nine sessions, each containing 18 trials, were run. The AV and VA (or VV1 and VV2) conditions were pseudorandomly presented within the session. To focus on the learning and learned phases of the task separately, the nine sessions were divided into the first six sessions (the learning phase with feedback) and the subsequent three sessions (the learned phase without feedback). The learning phase contained a total of 108 trials [54 AV (or VV1) trials and 54 VA (or VV2) trials].

MRI data acquisition. All images were acquired using a 3T MR scanner (Allegra; Siemens, Erlangen, Germany). For functional imaging during the sessions, an interleaved T2*-weighted gradient-echo echoplanar imaging (EPI) procedure was used to produce 34 continuous 4-mm-thick transaxial slices covering the entire cerebrum and cerebellum [repetition time (TR), 2000 ms; echo time (TE), 30 ms; flip angle, 75°; field of view, 192 mm; 64 × 64 matrix; voxel dimensions, 3.0 × 3.0 × 4.0 mm]. Oblique scanning was used to exclude the eyeballs from the images. The onset of each trial, relative to the preceding image acquisition, was jittered in steps of 500 ms within 1 TR (2000 ms) during the seventh to ninth sessions (learned phase), whereas there was no jittering in the first to sixth sessions (learning phase) (Dale, 1999). For anatomical imaging, T1-weighted magnetization-prepared rapid-acquisition gradient-echo (MP-RAGE) images, scanned at the same locations as those used for the EPI, were obtained for each subject [TR, 1460 ms; TE, 4.38 ms; flip angle, 8°; field of view, 192 mm (one slab); distant factor, 50%; number of slices per slab, 36; voxel dimensions, 0.9 × 0.8 × 4.0 mm]. To acquire a fine structural whole-head image, MP-RAGE images were also obtained (TR, 2500 ms; TE, 4.38 ms; flip angle, 8°; field of view, 230 mm (one slab); distant factor, 50%; number of slices per slab, 192; voxel dimensions, 0.9 × 0.9 × 1.0 mm).

Each session consisted of a continuous series of 365 vol acquisitions with a total duration of 12 min 14 s. To avoid subject fatigue, several breaks (of ~10 min) were inserted within the nine sessions (that is, in a typical case, three sessions/break/three sessions/break/three sessions). The total duration of the experiment was ~180 min, including the acquisition of the structural MR images.

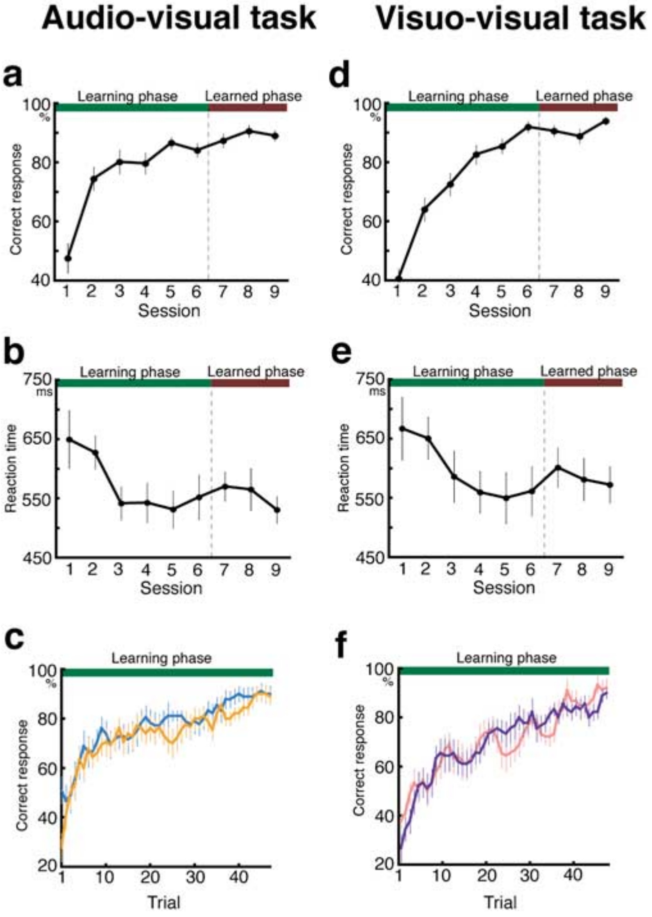

Figure 3. Plots of the behavioral data. ***a***, Time course of correct responses in the audiovisual task. The value in each session indicates the mean group performance for a session. ***b***, Time course of reaction times in the audiovisual task. The values for each session indicate the mean group reaction times for a session. ***c***, Moving average of six trials of correct responses in the AV (pale blue) or VA (pale orange) conditions in the audiovisual task. ***d***, Time course of correct responses in the visuo-visual task. ***e***, Time course of reaction times in the visuo-visual task. ***f***, Moving average of six trials of correct responses in VV1 (violet) or VV2 (rose) condition in the visuo-visual task. Error bars indicate SEM.

Image preprocessing. The first 7 vol of each session were eliminated to allow for the stabilization of the magnetization, and the remaining 358 vol per session (a total of 3222 vol per participant for nine sessions) were used for analysis. The data were preprocessed using Statistical Parametric Mapping 99 (SPM99) (Wellcome Department of Cognitive Neurology, London, UK). After correcting for differences in slice timing within each image volume, all volumes were realigned for motion correction. The same-slice position structural image volume was coregistered with the image volume of the eighth scan, and the whole-head MP-RAGE image volume was coregistered with this structural image volume. The whole-head image volume was normalized to the Montréal Neurological Institute T1 image template (Evans et al., 1994) using a nonlinear basis function. The same parameters were applied to all EPI volumes. The EPI volumes were spatially smoothed in three dimensions using a 10 mm full-width half-maximum Gaussian kernel.

Evaluation of the learning effects. To investigate the learning effects within the stimulus-related neural activity, we conducted regression analysis using the six learning phase sessions. The underlying idea is that the change in the stimulus-locked neural responses across the trials represents the learning effect. A schematic of this regression analysis is shown in Figure 2 and described below. We analyzed the AV (or VV1) and VA (or VV2) conditions separately.

For the AV condition, the MR signal data were first filtered with lowpass (4 mm Gaussian) and high-pass (cutoff frequency at 120 s) filters within each session.

AV (1, 1) is the scan volume acquired just before the initial presentation of the auditory S1, and AV (*i*, 1) is the *i*th presentation of S1. The *i*th AV-condition trial consists of AV (*i*, 1), AV (*i*, 2), . . . AV (*i*, 20), which represent the consecutive scan volumes acquired with a time interval of 2 s. In general, AV (*i*, *j*, *k*) represents the blood oxygen level-dependent signal of the *k*th voxel of the *j*th volume of the *i*th AV trial. Initially, within each trial, (1) percentage normalization and (2) linear detrending were performed.

AV $(i, \bar{j}, k)$ is the percentage of signal increase in the *k*th voxel of the *j*th scan volume in the *i*th AV trial compared with the baseline (average of the first two volume scan points, $j = 1, 2$) of the same voxel of the same trial:

$$\mathrm{AV}\,(i, \bar{j}, k) = 100 \; \frac{\mathrm{AV}\,(i, j, k) - \dfrac{\mathrm{AV}\,(i, 1, k) + \mathrm{AV}\,(i, 2, k)}{2}}{\dfrac{\mathrm{AV}\,(i, 1, k) + \mathrm{AV}\,(i, 2, k)}{2}} \qquad (1)$$

$(i = 1, \ldots, 54;\ j = 1, \ldots, 20)$.

Linear detrending within each trial was as follows:

$$\mathrm{AV}\,(i, \bar{\bar{j}}, k) = \mathrm{AV}\,(i, \bar{j}, k) - \{X(X'X)^{-1}X'\}\mathrm{AV}\,(i, \bar{j}, k)\,, \qquad (2)$$

where X is a 20×2 design matrix of a linear regression, and X' denotes the transposition of X.

The linear trend across the trials was evaluated in the *k*th voxel of the *j*th scan volume with a general linear model:

$$\mathrm{AV}\,(i, \bar{\bar{j}}, k) = \beta\, i + C + \varepsilon \quad (i = 1, \ldots, 54)\,, \qquad (3)$$

where β is the regression coefficient derived from the general linear model of AV $(i, \bar{\bar{j}}, k)$ on *i*, *C* is the effect of the *k*th voxel of the *j*th scan volume, and ϵ is the statistical error. As Equation 3 indicates, the model was applied without treatment of the intersession interval. This is based on the assumption that learning will not

→

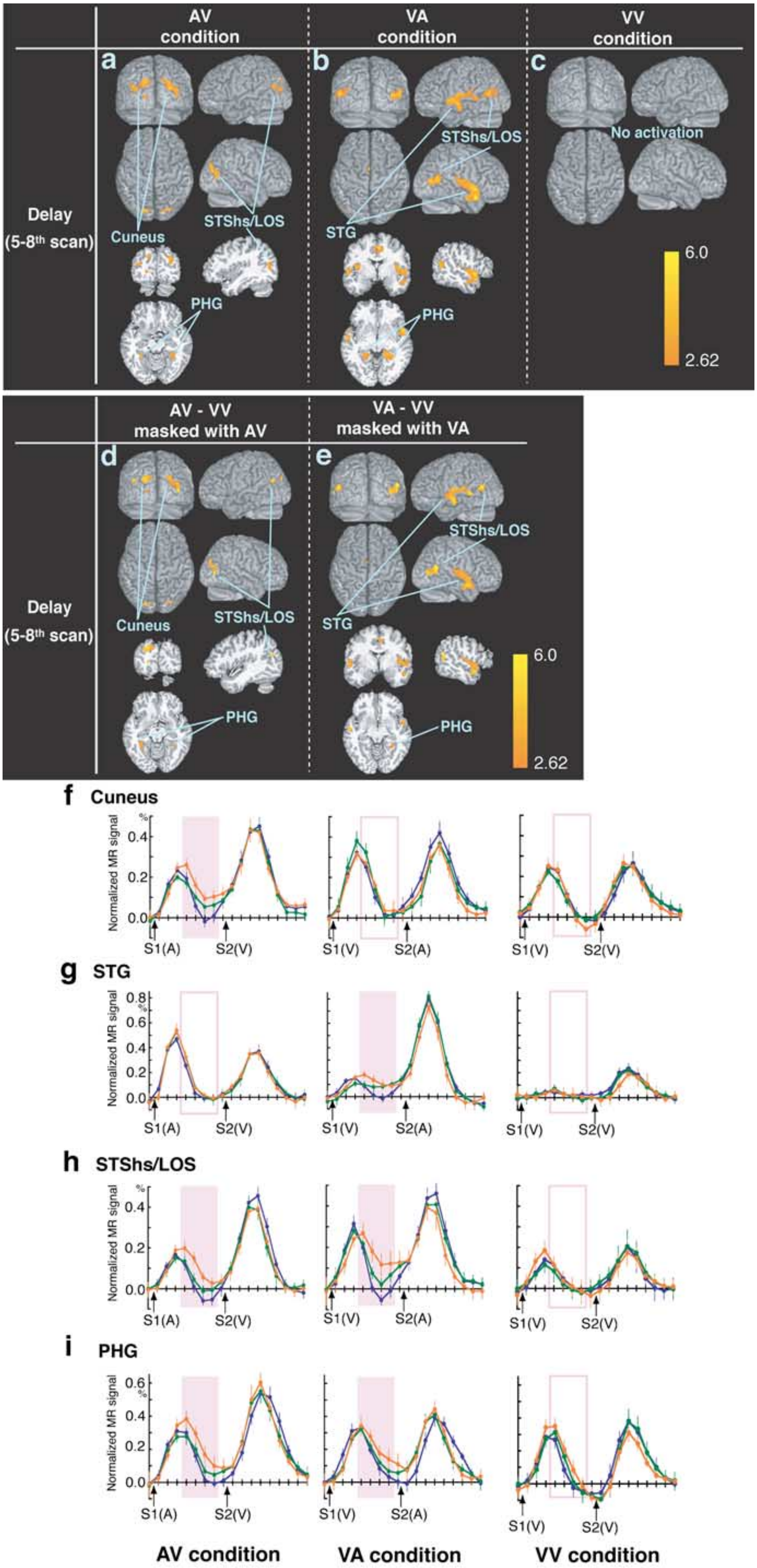

Figure 4. Regions showing changes in brain activity during the delay period in the AV (***a***; left column), VA (***b***; middle column), and VV (***c***; right column) conditions. Direct comparisons were also performed between the AV and VV (***d***) and VA and VV (***e***) conditions. Regions with positive (orange) changes are superimposed on surface-rendered or cross-sectioned high-resolution MRI scans. Scales show the *Z* values. ***f–i***, The time course of activation across trials in sessions 1 (blue), 4 (green), and 6 (orange). The percentage of normalized data was averaged within the sessions. Left, AV condition; middle, VA condition; right, VV condition. ***f***, Cuneus ($x = -20, y = -90, z = 22$). ***g***, STG ($x = -52, y = -28, z = 14$). ***h***, STShs/LOS ($x = 54, y = -64, z = 12$). ***i***, PHG ($x = 26, y = -42, z = -10$). Error bars indicate SEM.

Table 1. Regions showing positive changes in brain activity during the delay period as learning proceeded (corresponding to Fig. 4)

Condition	Cluster size (voxel number)	MNI coordinate			Z value	Side	Location	BA
		x	*y*	*z*				
AV	113	−20	−90	20	3.44	L	Cuneus	18/19
	337	28	−86	28	3.19	R	Cuneus	19
		40	−72	14	3.28	R	STShs/LOS	39/19
	64	−40	−72	22	2.92	L	STShs/LOS	39/19
	437	−30	−40	−16	3.14	L	PHG	36
	124	36	−40	−18	3.18	R	PHG	36
	53	−12	−42	0	2.90	L	PHG	36
VA	1607	56	10	−4	3.74	R	STG	22
		60	0	4	3.05	R	MTG	21
		64	−18	12	2.80	R	TTG	42
		42	2	10	3.04	R	Insula	13
	1131	−66	−28	16	3.19	L	STG	22
		−58	−10	−8	3.16	L	MTG	21
		−60	−10	14	3.16	L	TTG	42
		−40	−8	12	3.34	L	Insula	13
	334	54	−64	12	3.38	R	STShs/LOS	39/19/22
	114	−56	−66	12	3.31	L	STShs/LOS	39/19/22
	421	26	−42	−10	3.66	R	PHG	36
	289	−20	−58	−4	2.99	L	PHG	36
	115	34	−16	−14	2.83	R	PHG/HC	36
VV	NS							

BA, Brodmann's area; HC, hippocampus; MNI, Montréal Neurological Institute; MTG, middle temporal gyrus; TTG, transverse temporal gyrus; NS, not significant; L, left; R, right.

progress during the intersession interval. Error trials were disregarded, because we were not able to discriminate between an erroneous button press and an incorrect answer (which is not an error). A contrast image that contained the slope (β) estimate of every voxel was generated at each scan point in each individual. Therefore, 20 contrast images were obtained from each subject, because each trial contained 20 scan points. Within each trial, S1 was presented between $j = 1$ and 2, S2 was presented between $j = 9$ and 10, and F was presented between $j = 10$ and 11. The peak of the MR signal change in response to S1 was around $j = 4$, attributable to delayed hemodynamic response. Hence, the MRI signal at $j = 5$, 6, 7, and 8 should reflect the neural activity of the earlier delay period, in addition to the neural activity in response to the presentation of S1. To explore and summarize the continuous increase (positive slope) or decrease (negative slope) in MRI signal during the delay period, contrast images of $j = 5$, 6, 7, and 8 were averaged for each subject. Similarly, contrast images of $j = 12$, 13, and 14 corresponding to the S2/F-related response were averaged for each subject. Group inference was evaluated by one-sample *t* tests using the averaged contrast images of each subject, which represent the evolving crossmodal responses during the delay period or the S2/F-related response (S2/feedback). An identical procedure was applied for VA, VV1, and VV2.

To evaluate whether the crossmodal response during the delay period was specific to the crossmodal trials, two-sample *t* tests were conducted to directly compare AV with VV or VA with VV. An inclusive mask method was used to confirm the areas in which there were signal increases in the AV or VA condition. The statistical thresholds were set at $z > 2.33$, and the cluster size was set at >50 voxels.

Evaluation of sustained activity during delay period without learning effect. To detect sustained activity during the delay period that corresponded to the working memory component of the delayed matching-to-sample task, the data were analyzed using a conventional SPM approach. This is because, without the learning effect, each trial could be regarded as a repetition. Hence, SPM is the most powerful method of depicting the event-related activation for S1, the delay period, and S2/F. To show the neural substrates of the task without the learning effects, in each subject, we modeled the transient neural responses to S1 and S2/F, as well as the sustained activity between S1 and S2. Contrast images of the sustained activity of each subject were used for the group analysis with a random-effects model to obtain population inferences (Friston et al., 1999). The resulting set of voxel values for each contrast constituted a statistical parametric map of the *t* statistic (SPM{*t*}), which was transformed to the normal distribution unit (SPM{*Z*}). The threshold for SPM{*Z*} was set at $Z > 3.09$ and $p < 0.05$ with a correction for multiple comparisons at the cluster level for the entire brain (Friston et al., 1996). To evaluate the laterality of the activation patterns, two-sample *t* tests were also conducted to compare the original images with flipped contrast images. The threshold for SPM{*Z*} was set at $Z > 3.09$. An identical procedure was applied for the VA, VV1, and VV2 conditions.

Results

Performance

During the first six sessions (learning phase) of the audiovisual paired-association task, the proportion of correct responses made by the 15 subjects increased as the sessions proceeded (Fig. 3*a*) (repeated-measures ANOVA; $F_{(3.29,49.30)} = 28.54$; $p < 0.001$ with Greenhouse-Geisser correction), and reaction times decreased (Fig. 3*b*) (repeated-measures ANOVA; $F_{(2.88,43.13)} = 4.67$; $p < 0.005$ with Greenhouse-Geisser correction). There was no difference in learning speed between the AV and VA stimulus-order trial conditions, as shown in Figure 3*c* (six trials; moving averaged data).

In contrast to the learning phase, there were no statistically significant changes in the accuracy and reaction times during the final three sessions [learned phase; repeated-measures ANOVA; $F_{(2,30)} = 2.13$ (not significant) for correct responses; $F_{(2,30)} = 0.95$ (not significant) for reaction time] (Fig. 3*a*,*b*). Hence, the subjects had learned the arbitrary audiovisual associates during the preceding (learning) sessions. According to the answer given in the subject's debriefing after the experiment, as learning proceeded, the paired associate was triggered by the S1 presentation. A similar pattern was seen in the visuo-visual paired-association task completed by another group of 15 subjects (Fig. 3*d*–*f*). There was no clear difference in learning speed between the audiovisual and visuo-visual paired-association tasks. There was no difference in learning speed between the VV1 and VV2 conditions, as shown in Figure 3*f* (six trials; moving averaged data). Therefore, for additional imaging analysis, the results of the VV1 condition were used as those of VV stimulus-order trial condition to compare them with the AV and VA conditions.

Evaluation of learning effects

During the delay period of the audiovisual paired-association task, the response to the auditory S1 was gradually enhanced in the visual cortex (cuneus) (Fig. 4*a*,*f*, Table 1), and the response to the visual S1 increased in the superior temporal gyrus (STG) (Fig. 4*b*,*g*, Table 1). The junction between the horizontal posterior segment of the STS (STShs) and the lateral occipital sulcus (LOS), designated as the STShs/LOS, and the parahippocampal gyrus (PHG) revealed a gradual increase in signal in response to both auditory (Fig. 4*a*,*h*,*i*, Table 1) and visual S1s (Fig. 4*b*,*h*,*i*, Table 1) during the delay period. No decrements in activity were observed. In contrast to the audiovisual association task, there was no obvious signal change during the delay period in the visuo-visual association task (Fig. 4*c*,*f*–*i*). A direct comparison of AV–VV revealed activation patterns similar to those during the AV condition, and VA–VV was the same as VA, confirming that the neural activation underlying the learning effect is specific to the crossmodal association task (Fig. 4*d*,*e*). During the AV condition, the delay-period activities in the cuneus, STShs/LOS, and PHG were significantly correlated with task performance. During the VA condition, the activity levels in the STG, STShs/LOS, and PHG were significantly correlated. During the VV condition, no such correlation was found (supplemental Fig. 1, available at www.jneurosci.org as supplemental material).

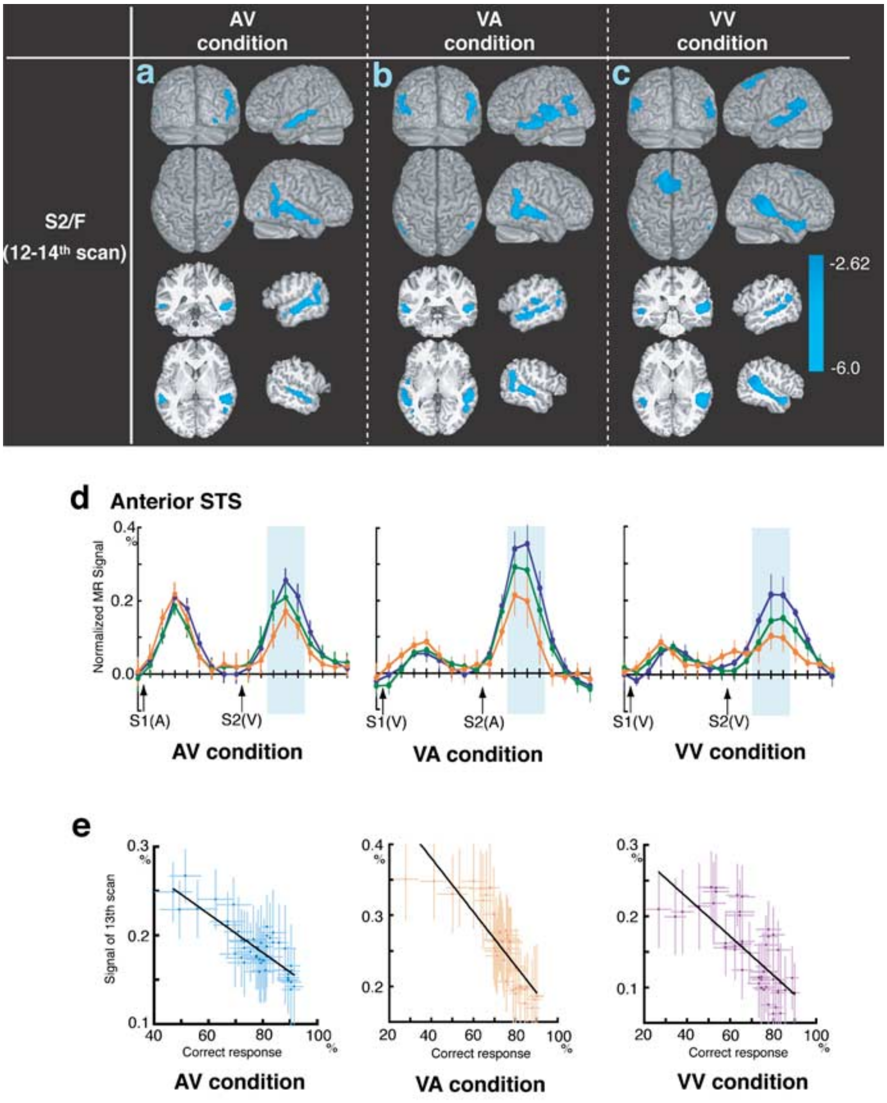

Figure 5. ***a*–*c***, Regions showing changes in brain activity after the S2/F period in the AV (***a***), VA (***b***), and VV (***c***) conditions. Regions with negative (sky blue) changes are superimposed on surface-rendered or cross-sectioned high-resolution MRI scans. Scales show the *Z* values. ***d***, The time courses of the STS ($x = -60, y = -22, z = -4$) activation across trials in sessions 1 (blue), 4 (green), and 6 (orange). The percentage of normalized data was averaged within the session. Left, AV condition; middle, VA condition; right, VV condition. The closed sky-blue column indicates a significant decrease in the percentage of normalized signal. Error bars indicate SEM. ***e***, Correlations between the performance ratio and signals in the STS. Six-trial moving averaged performance ratios and signals of the 13th scan in the STS ($x = -60, y = -22, z = -4$) were used. The correlation coefficients (r^2) of the AV, VA, and VV conditions were 0.69, 0.74, and 0.56, respectively. All coefficients were statistically significant ($p < 0.01$).

No increments in S2/F-related activities were observed in either the audiovisual or visuo-visual association tasks. Decreasing activity levels were found in the area along the STS in the AV, VA, and VV conditions (Fig. 5*a*–*c*, Table 2). The presupplementary motor area (preSMA) also showed decreased activity in the VV condition (Fig. 5*c*, Table 2). The polymodal STS was activated by the S2/F stimuli; this activation was maximal in the first session and decreased as learning proceeded (Fig. 5*d*). Signal values for the S1 peak (fourth scan) remained unchanged throughout the AV, VA, and VV conditions. In contrast, the signal values for the S2/F peaks (13th scan) were maximal during the initial learning phase and decreased as learning proceeded (Fig. 5*d*). To clarify the relationship between performance and the STS signal, the correlation between correct responses and signal values on the 13th scan was plotted and fitted to a linear approximation. The results showed that the task-related signal changes in the STS were significantly correlated with task performance in all three trial conditions (Fig. 5*e*).

Sustained activity in the delay period

The intraparietal sulcus (IPS), the preSMA, the dorsal part of the premotor area, the inferior frontal gyrus, and the lateral prefrontal cortex (LPFC) were activated constantly throughout the delay period in the AV, VA, and VV conditions (Fig. 6*a*–*c*, Table 3). The subtraction of the flipped activation images from the originals confirmed that the activity was strongly left lateralized in all three trial conditions (Fig. 6*d*–*f*). A time course plot of the averaged signals of those regions in the first, fourth, and sixth sessions revealed sustained activation during the learning phase (Fig. 6*g*–*j*), with the exception of the preSMA in the AV condition, in which the signal increased after sustained activity during the late delay period (Fig. 6*k*).

Discussion

Crossmodal activation

As expected, the auditory S1 activated visual areas and the visual S1 activated the auditory cortex as learning proceeded. This is consistent with nonhuman primate studies showing crossmodal activation of neurons in the auditory association cortex in re-

Table 2. Regions showing negative changes in brain activity after S2/F as learning proceeded (corresponding to Fig. 5)

Condition	Cluster size (voxel number)	MNI coordinate			Z value	Side	Location	BA
		x	y	z				
AV	1849	54	−20	−8	4.17	R	STS	21/22
		58	−42	2	3.99	R	STS	21/22
		56	−52	24	3.62	R	SMG	40
	642	−60	−22	−4	3.91	L	STS	21/22
		−60	−34	4	3.46	L	STS	21/22
VA	1632	56	−20	−4	3.26	R	STS	21/22
		54	−38	2	4.20	R	STS	21/22
		60	−52	18	4.07	R	SMG	40
	1619	−62	−24	−2	3.13	L	STS	21/22
		−64	−42	2	3.76	L	STS	21/22
	382	−60	−58	28	3.21	L	SMG	40
VV	1154	54	−16	−8	2.92	R	STS	21/22
		48	−36	0	3.98	R	STS	21/22
	776	−52	−22	−6	3.39	L	STS	21/22
		−54	−42	10	3.37	L	STS	21/22
	1084	−2	16	62	3.44	L	preSMA	6

BA, Brodmann's area; MNI, Montréal Neurological Institute; NS, not significant; SMG, supramarginal gyrus; STS, superior temporal sulcus; L, left; R, right.

sponse to somatosensory (Fu et al., 2003) or visual stimuli (Schroeder and Foxe, 2002). Using visuo-tactile delayed matching-to-sample tasks, Zhou and Fuster (1996, 1997, 2000) demonstrated that somatosensory neurons reacted to visual stimuli associated with tactile sensations, and some showed sustained activation during delay periods. The crossmodal activation of unimodal areas by anticipatory paired associates suggests that the sensory memory of a particular modality is stored in parasensory association cortex (Fuster, 1997). The more vivid the memory, the more prominent the reactivation in these regions (Wheeler et al., 2000).

The STShs/LOS and PHG responses to S1 were augmented throughout learning, regardless of the modality of S1. This indicates that crossmodal responses in unimodal areas might be mediated through polymodal association and memory-related regions. Crossmodal responses are observed in the STShs, a posterior polysensory extension of the STS (Calvert, 2001; Poremba et al., 2003; Beauchamp et al., 2004). Because the STShs was activated by S1 regardless of stimulus modality, this polymodal region might be linked to unimodal regions to represent the paired associates. The medial temporal lobe may be important in the reactivation of long-term memories (Squire and Zola, 1997; Nyberg et al., 2000). Ranganath and D'Esposito (2001) and Sakai et al. (2002) demonstrated PHG activation during stimulus retrieval even in short-term memory tasks. Hence, the increased PHG activity during learning probably reflects the retrieval of paired associates during the delay period, in preparation for the second stimulus. Because this pattern was not observed during visuo-visual intramodal association learning, the crossmodal associations might be specifically formed through "binding" unimodal sensory areas via polymodal regions.

According to Fuster et al. (2000), the matching of visual and auditory stimuli across time involves the following: (1) activation of the network representing the crossmodal association in permanent storage, (2) sustained activation of that association in working memory, and (3) reactivation of the network during the presentation of the paired associates. Here, the term binding indicates the formation of a network representing the crossmodal association. This is based on the crossmodal activation during the delay period, which may represent the paired patterns evoked by S1 presentation. Hence, the major difference between crossmodal matching and crossmodal integration (Stein and Meredith, 1993) is that the former involves associating visual and auditory stimuli across time, whereas the latter requires simultaneous presentation of different modalities.

STS for binding stimuli

The S2/F-related activation along the STS was high during initial learning and decreased as learning proceeded, implying that these areas were involved in creating the association between the stimuli. The superior temporal polysensory (STP) cortex, the homolog of the human STS, is a polymodal area in nonhuman primates (Benevento et al., 1977; Poremba et al., 2003). Because the STP cortex is connected to unimodal visual and auditory areas, as well as the amodal medial temporal lobe and prefrontal cortex (Blatt et al., 2003; Padberg et al., 2003), the human STS is well suited to the formation of both crossmodal and intramodal linkages. Although it was suggested that the STS is where auditory and visual information about objects is integrated (Beauchamp et al., 2004), it is not clear how the STS establishes the association between arbitrary visual and auditory stimuli. Our subjects were required to link two arbitrary temporally separated stimuli on the basis of feedback information. Previous neuroimaging studies of crossmodal learning without responses or feedback showed no activation in the STS/middle temporal gyrus (McIntosh et al., 1998; Gonzalo et al., 2000), suggesting that STS activation is feedback related. Significant negative correlations between performance and S2/F-related signal increments in the STS (Fig. 5*e*) also suggest that the STS activation is related to the learning workload: the arbitrary relationship between the two stimuli gradually becomes related as learning proceeds; therefore, less work is required to link the two stimuli based on feedback. In contrast to the delay-period activity difference seen between audiovisual and visuo-visual association learning, decreased STS activity after the S2/F stimulus was observed in both tasks (Fig. 5*a*–*c*). Thus, the STS might be involved in building the paired association, regardless of the modality of the stimuli.

Although the STS is known to be involved in top-down attentional control (Hopfinger et al., 2000), the decrement of the S2/F response in the STS is unlikely the result of a decline of general attention. The STS is the only area that showed a decrease in the S2/F response, whereas the response to S1 was unchanged throughout learning (Fig. 5). Were the decrement in activation in the STS attributable to decreasing attention, other areas that are part of the attentional network would also show decreased activation. In addition, were the STS involved in general attentional control during a trial, the signals would have declined in response to S1 as well as to S2/F.

Another possible explanation is that there was a reduction in error monitoring between the early and late learning phases. The need for error monitoring gradually decreases as learning proceeds. Because it was not possible to separate the feedback from S2 in the present study, we could not completely exclude this explanation for the decreased activity in the STS. However, error detection could be attributed to the anterior cingulate cortex. Furthermore, there are no previous reports of STS involvement in error monitoring (Ullsperger and von Cramon, 2003; Holroyd et al., 2004). Therefore, it is unlikely that the decrement is attributable to a decreasing demand for error monitoring. Together, the present results support the hypothesis that STS is an associative learning device, or an area that links different types of information regardless of the stimulus modality, even across visual and auditory modalities (Beauchamp et al., 2004).

Working memory

Left-lateralized areas, such as the LPFC, premotor areas and the IPS, were active during the delay period and did not show a learning effect. This left-lateralized parietal-premotor-prefrontal network might be related to the maintenance of information in an on-line state for a brief period of time, so-called working memory (Baddeley, 1992; Smith and Jonides, 1998). This finding suggests a separation between the neural substrates involved in storing memories of paired associates and those subserving working memory. Interestingly, the present finding represents the left–right reversed activation pattern of the parietal-premotor-prefrontal network for spatial working memory (D'Esposito et al., 1998; Smith and Jonides, 1998; Wager and Smith, 2003). In our study, the information held during the delay period was not spatial but either visual textures or

→

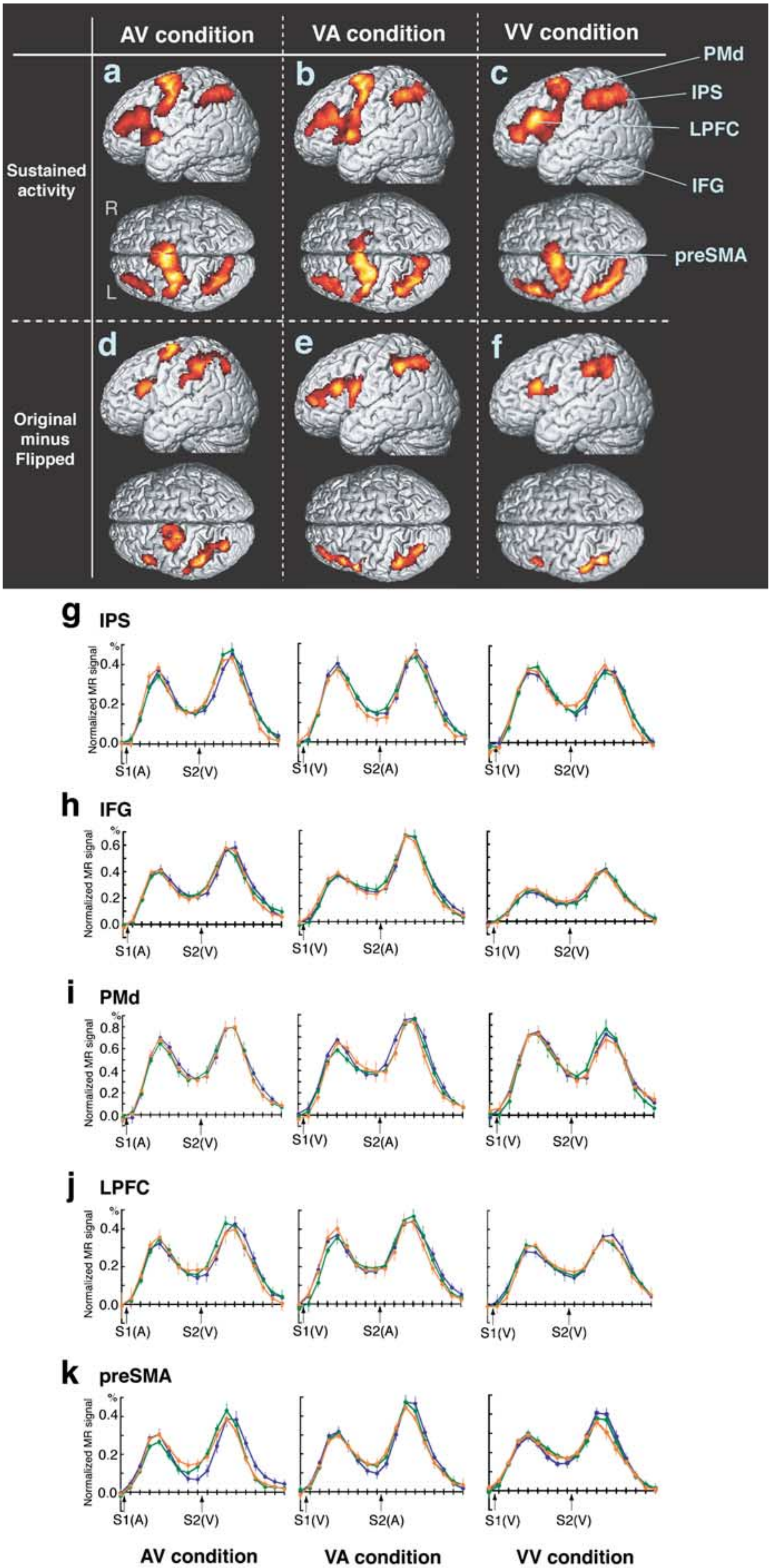

Figure 6. ***a–c***, Sustained activities during the delay period in the AV (***a***), VA (***b***), and VV (***c***) conditions are superimposed on surface-rendered high-resolution MRI scans. ***d–f***, Comparison between the original and right–left flipped contrast images revealed left-lateralized activation in the AV (***d***), VA (***e***), and VV (***f***) conditions. The time course of activation collapsed across trials in sessions 1 (blue), 4 (green), and 6 (orange) in IPS (***g***; $x = -28, y = -58, z = 40$), IFG (***h***; $x = -46, y = 14, z = 2$), PMd (***i***; $x = -30, y = -2, z = 66$), LPFC (***j***; $x = -42, y = 42, z = 20$), and preSMA (***k***; $x = -6, y = 8, z = 64$). IFG, Inferior frontal gyrus; PMd, dorsal part of premotor area. Error bars indicate SEM.

Table 3. Regions showing sustained brain activity during the delay period throughout the learning process (corresponding to Fig. 6)

Condition	Cluster size (voxel number)	MNI coordinate			*Z* value	Side	Location	BA
		x	*y*	*z*				
AV		−30	−2	66	4.97	L	PMd	6
		−6	8	64	4.93	L	preSMA	6
	2511	−46	14	2	3.78	L	IFG	45
		−42	32	26	3.38	L	LPFC	46/10
	3018	−30	−62	50	4.28	L	IPS	7/40
VA	6560	−22	−10	62	4.87	L	PMd	6
		8	2	62	3.49	L	preSMA	6
		−42	14	16	3.66	L	IFG	44/45
		−42	42	20	4.03	L	LPFC	46/10
	2222	−28	−58	40	4.87	L	IPS	7/40
VV	6410	−30	−6	64	3.87	L	PMd	6
		−12	6	60	5.53	L	preSMA	6
		−44	14	8	4.83	L	IFG	44/45
		−46	30	26	5.93	L	LPFC	46/10
	3678	−34	−58	46	5.24	L	IPS	7/40

BA, Brodmann's area; IFG, inferior frontal gyrus; MNI, Montréal Neurological Institute; PMd, dorsal part of premotor area; L, left; R, right.

amplitude-modulated sounds. The left-lateralized activation is partly consistent with the idea of Manoach et al. (2004) that auditory working memory is stored in the left frontoparietal network; however, we observed strong left-lateralized activation even in the visuo-visual intramodal association task. This suggests that the left frontoparietal network was driven even by nonspatial items. Verbalization of the stimuli was unlikely, because verbal working memory involves the inferior parietal lobule rather than the IPS (Smith and Jonides, 1998; Gruber and von Cramon, 2003; Veltman et al., 2003).

The signal change in the preSMA during learning might be intermingled with two types of activity. The first is a sustained activity during the delay period, which is unchanged throughout learning. Because the preSMA is involved in the parietal-premotor-prefrontal network supporting working memory function (Hartley and Speer, 2000), this sustained activation represents the working memory component. The second type of activity is a signal change during the late delay period that corresponds to learning. As learning proceeded, the linkage between the stimulus and response (button press) was strengthened; the subjects could wait and prepare the response because they had already recognized the pair, even if they did not know which button to press until the presentation of the second stimulus. Therefore, the enhanced preSMA activation might not reflect a simple readiness to respond, but rather the content of the response. The premotor area might play a role in the readiness to respond as well as working memory function.

Absence of prefrontal activation

We did not observe learning-associated signal changes in the prefrontal cortex, although in nonhuman primates it was shown to be involved in top-down voluntary recall (Hasegawa et al., 1998; Tomita et al., 1999) and crossmodal association (Fuster et al., 2000). This may be attributable to the difference in the duration of the learning period. In previous studies, the monkeys recalled items from long-term memory, whereas the subjects in the present study recalled recently memorized items. Additional research is required to clarify this issue.

Conclusion

In summary, the representation of crossmodal paired associates includes unimodal visual and auditory areas activated by the visual and auditory associates, respectively. The arbitrary crossmodal association might be accomplished by binding the unimodal sensory areas through occipitotemporal association and memory-related areas. In contrast, the STS might play a key role in building paired associations, regardless of the modality. The neural substrates of working memory are segregated from those for memory storage and formation.

References

Baddeley A (1992) Working memory. Science 255:556–559.

Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41:809–823.

Benevento LA, Fallon J, Davis BJ, Rezak M (1977) Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. Exp Neurol 57:849–872.

Blatt GJ, Pandya DN, Rosene DL (2003) Parcellation of cortical afferents to three distinct sectors in the parahippocampal gyrus of the rhesus monkey: an anatomical and neurophysiological study. J Comp Neurol 466:161–179.

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123.

Dale AM (1999) Optimal experimental design for event-related fMRI. Hum Brain Mapp 8:109–114.

D'Esposito M, Aguirre GK, Zarahn E, Ballard D, Shin RK, Lease J (1998) Functional MRI studies of spatial and nonspatial working memory. Brain Res Cogn Brain Res 7:1–13.

Evans AC, Kamber M, Collins DL, MacDonald D (1994) An MRI-based probalistic atlas of neuroanatomy. In: Magnetic resonance scanning and epilepsy (Shorvon SD, ed), pp 263–274. New York: Plenum.

Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD (1996) Detecting activations in PET and fMRI: levels of inference and power. NeuroImage 4:223–235.

Friston KJ, Holmes AP, Worsley KJ (1999) How many subjects constitute a study? NeuroImage 10:1–5.

Fu KM, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE (2003) Auditory cortical neurons respond to somatosensory stimulation. J Neurosci 23:7510–7515.

Fuster JM (1997) Network memory. Trends Neurosci 20:451–459.

Fuster JM, Bodner M, Kroger JK (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. Nature 405:347–351.

Gibson JR, Maunsell JH (1997) Sensory modality specificity of neural activity related to memory in visual cortex. J Neurophysiol 78:1263–1275.

Gonzalo D, Shallice T, Dolan R (2000) Time-dependent changes in learning audiovisual associations: a single-trial fMRI study. NeuroImage 11:243–255.

Gruber O, von Cramon DY (2003) The functional neuroanatomy of human working memory revisited. Evidence from 3-T fMRI studies using classical domain-specific interference tasks. NeuroImage 19:797–809.

Hartley AA, Speer NK (2000) Locating and fractionating working memory using functional neuroimaging: storage, maintenance, and executive functions. Microsc Res Tech 51:45–53.

Hasegawa I, Fukushima T, Ihara T, Miyashita Y (1998) Callosal window between prefrontal cortices: cognitive interaction to retrieve long-term memory. Science 281:814–818.

Holroyd CB, Nieuwenhuis S, Yeung N, Nystrom L, Mars RB, Coles MG, Cohen JD (2004) Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. Nat Neurosci 7:497–498.

Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. Nat Neurosci 3:284–291.

Manoach DS, White NS, Lindgren KA, Heckers S, Coleman MJ, Dubal S, Holzman PS (2004) Hemispheric specialization of the lateral prefrontal cortex for strategic processing during spatial and shape working memory. NeuroImage 21:894–903.

McIntosh AR, Cabeza RE, Lobaugh NJ (1998) Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. J Neurophysiol 80:2790–2796.

Miyashita Y (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature 335:817–820.

Miyashita Y (2000) Visual associative long-term memory: encoding and retrieval in inferotemporal cortex of the primate. In: The new cognitive neuroscience, Ed 2 (Gazzaniga MS, ed), pp 379–392. Cambridge, MA: MIT.

Miyashita Y, Hayashi T (2000) Neural representation of visual objects: encoding and top-down activation. Curr Opin Neurobiol 10:187–194.

Nyberg L, Habib R, McIntosh AR, Tulving E (2000) Reactivation of encoding-related brain activity during memory retrieval. Proc Natl Acad Sci USA 97:11120–11124.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113.

Padberg J, Seltzer B, Cusick CG (2003) Architectonics and cortical connections of the upper bank of the superior temporal sulcus in the rhesus monkey: an analysis in the tangential plane. J Comp Neurol 467:418–434.

Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev 87:532–552.

Poremba A, Saunders RC, Crane AM, Cook M, Sokoloff L, Mishkin M (2003) Functional mapping of the primate auditory system. Science 299:568–572.

Ranganath C, D'Esposito M (2001) Medial temporal lobe activity associated with active maintenance of novel information. Neuron 31:865–873.

Rescorla RA, Wagner AR (1972) A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical conditioning II: current research and theory (Black AH, Prokasy WF, eds), pp 64–99. New York: Appleton Century Crofts.

Sakai K, Miyashita Y (1991) Neural organization for the long-term memory of paired associates. Nature 354:152–155.

Sakai K, Rowe JB, Passingham RE (2002) Parahippocampal reactivation signal at retrieval after interruption of rehearsal. J Neurosci 22:6315–6320.

Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Brain Res Cogn Brain Res 14:187–198.

Smith EE, Jonides J (1998) Neuroimaging analyses of human working memory. Proc Natl Acad Sci USA 95:12061–12068.

Squire LR, Zola SM (1997) Amnesia, memory and brain systems. Philos Trans R Soc Lond B Biol Sci 352:1663–1673.

Stein BE, Meredith MA (1993) The merging of the senses, pp 172–173. Boston: MIT.

Tomita H, Ohbayashi M, Nakahara K, Hasegawa I, Miyashita Y (1999) Top-down signal from prefrontal cortex in executive control of memory retrieval. Nature 401:699–703.

Ullsperger M, von Cramon DY (2003) Error monitoring using external feedback: specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. J Neurosci 23:4308–4314.

Veltman DJ, Rombouts SA, Dolan RJ (2003) Maintenance versus manipulation in verbal working memory revisited: an fMRI study. NeuroImage 18:247–256.

Wager TD, Smith EE (2003) Neuroimaging studies of working memory: a meta-analysis. Cogn Affect Behav Neurosci 3:255–274.

Wheeler ME, Petersen SE, Buckner RL (2000) Memory's echo: vivid remembering reactivates sensory-specific cortex. Proc Natl Acad Sci USA 97:11125–11129.

Zhou YD, Fuster JM (1996) Mnemonic neuronal activity in somatosensory cortex. Proc Natl Acad Sci USA 93:10533–10537.

Zhou YD, Fuster JM (1997) Neuronal activity of somatosensory cortex in a cross-modal (visuo-haptic) memory task. Exp Brain Res 116:551–555.

Zhou YD, Fuster JM (2000) Visuo-tactile cross-modal associations in cortical somatosensory cells. Proc Natl Acad Sci USA 97:9777–9782.