Cross-Modal Temporal Order Memory for Auditory Digits and Visual Locations: An fMRI Study

Daren Zhang,^{1*} Xiaochu Zhang,¹ Xiwen Sun,^{1,2} Zhihao Li,¹
Zhaoxin Wang,¹ Sheng He,³ and Xiaoping Hu⁴

¹Department of Neurobiology and Biophysics, University of Science and Technology of China, Hefei, Anhui, People's Republic of China

²Department of MRI, Huadong Hospital, Shanghai, People's Republic of China

³Department of Psychology, University of Minnesota, Minneapolis, Minnesota

⁴Department BME, Emory University and Georgia Tech, Atlanta, Georgia

Abstract: A function of working memory is to remember the temporal sequence of events, often occurring across different sensory modalities. To study the neural correlates of this function, we conducted an event-related functional magnetic resonance imaging (fMRI) experiment with a cross-modal memory task. Subjects were required to recall auditory digits and visual locations either in mixed order (cross-modality) or in separate order (within-modality). To identify the brain regions involved in the memory of cross-modal temporal order, we compared the blood oxygenation level-dependent (BOLD) response between the mixed and the separate order tasks. As a control, cortical areas sensitive to the memory load were mapped by comparing the 10-item condition with the 6-item condition in the separate order task. Results show that the bilateral prefrontal, right premotor, temporo-parietal junction (TPJ) and left superior parietal cortices had significantly more activation in the mixed task than in the separate task. Some of these areas were also sensitive to the memory load, whereas the right prefrontal cortex and TPJ were relatively more sensitive to the cross-modal order but not the memory load. Our study provides potential neural correlates for the episodic buffer, a key component of working memory as proposed previously [Baddeley. *Trends Cogn Sci* 2000;4:417–423]. *Hum. Brain Mapp.* 22:280–289, 2004. © 2004 Wiley-Liss, Inc.

Key words: working memory; episodic buffer; central executive; prefrontal cortex; active integration; modality effect

INTRODUCTION

Cross-modal integration and remembering temporal order are two important and basic cognitive functions. Humans are equipped with multiple sensory systems that are constantly registering input information. We can bind information that occurs in the temporal and the spatial domain. Binding can occur for information from the same spatial location over time, at the same time from multiple locations, or even across multiple locations over time. Furthermore, information does not have to come from the same sensory modality. With this integration, events that we remember are no longer isolated in their spatial and temporal context. For example, in a conference, a speaker's presentation is composed of her or his voice and the slides presented on the screen. One can integrate the voice (auditory modality) and slides (visual modality) in the temporal order of the presentation to get a meaningful story. It is an integration process

Contract grant sponsor: National Nature Science Foundation of China; Contract grant number: 39928005, 39970253, 30370478, 30328017; Contract grant sponsor: National Basic Research Program of China; Contract grant number: G1998030509; Contract grant sponsor: Chinese Academy of Sciences; Contract grant sponsor: NIH; Contract grant number: RO1EB002009; Contract grant sponsor: James S. McDonnell Foundation.

*Correspondence to: Dr. Daren Zhang, Department of Neurobiology and Biophysics, University of Science and Technology of China, Hefei, Anhui, 230026, P.R. China. E-mail: drzhang@ustc.edu.cn

Received for publication 12 October 2003; Accepted 12 February 2004

DOI: 10.1002/hbm.20036

Published online in Wiley InterScience (www.interscience.wiley.com).

that binds information distributed in temporal domain and allows the individual to remember the events occurring across different sensory modalities in their temporal order.

Cross-modal integration can occur at the perception, attention, and memory levels [Baddeley, 2000; Calvert et al., 2001; Corbetta and Shulman, 2002; Driver and Spence, 2000]. The present study focuses on the neural basis of temporal integration in cross-modal memory.

Both cross-modal processing and memory of temporal order have received increased attention in recent years. Neural imaging studies show that several cortical areas, mainly the prefrontal cortex (PFC) and the intraparietal lobule are involved in cross-modal integration [Badgaiyan et al., 2001; Bushara et al., 2003; Calvert et al., 2000; Downar et al., 2000; Klingberg and Roland, 1998; Laurienti et al., 2003], and the PFC and premotor cortex take part in the memory of temporal order [Chein and Fiez, 2001; Marshuetz et al., 2000; Rao et al., 2001; Rowe and Passingham, 2001].

Behavioral studies show a cross-modal order effect (i.e., modality effect) in recalling items from two modalities. Penney [1989] states, "When subjects are forced to recall items from a mixed-mode list according to a strict temporal order, recall is lowered relative to a condition in which subjects order items within each modality." The nature of the neural mechanism for memory of cross-modal temporal order remains an open question. In the classic working memory model, the two slave systems (phonological loop and visuospatial sketchpad) process and store verbal and visuospatial information separately [Baddeley, 1986, 1992]. This model works for separate memory tasks, but it does not work for cross-modal tasks such as recalling items in a mixed order from two modalities. Baddeley [2000] has pointed out that "...There are, however, a number of phenomena that are not readily captured by the original model." In the original model, the central executive "contained no short-term multimodal store capable of holding such complex representations." Consequently, an episodic buffer is proposed, which "comprises a limited capacity system that provides temporary storage of information held in a multimodal code, which is capable of binding information from the subsidiary systems ..." He further pointed out that "The buffer is episodic in the sense that it holds episodes whereby information is integrated across space and potentially extended across time" [Baddeley 2000]. This episodic buffer seems well suited for the memory of temporal orders for events registered across sensory modalities. The neural correlates of this buffer, however, have not been elucidated. Through investigating in the present study the neural correlates of remembering the cross-modal temporal order, we hope to shed some light to the neural mechanism of the episodic buffer.

We conducted an event-related functional magnetic resonance imaging (fMRI) experiment of cross-modal memory. Subjects recalled auditory digits and visuospatial locations in mixed (i.e., cross-modal) temporal order or separate (i.e., within-modal) temporal order. Because previous imaging studies have shown that the prefrontal cortices (PFCs) are

involved in both the integration of cross-modal information [Calvert et al., 2001; Downar et al., 2000] and temporal information processing or remembering of temporal orders [Bushara et al., 2001; Marshuetz et al., 2000; Rao et al., 2001], we were naturally interested in the role of PFC in the memory of cross-modal temporal order. Recalling in mixed order, however, is more difficult than recalling in separate order when the number of items recalled is the same. In this study, the task difficulty determines the mental load, which is reflected in the different levels of performance. The heavier load is thus a salient feature of the mixed recall, and PFCs have been shown to be sensitive to memory load or task difficulty [Rypma and D'Esposito, 1999]. Differentiating a specific function from an effect of general task difficulty was a common concern in many studies on PFC [Laurienti et al., 2003; Marshuetz et al., 2000; Rowe and Passingham, 2001]. To delineate the involvement of PFC in the integration of cross-modal order from difficulty of task or memory load, we also included a higher load condition in the separate order task.

To identify brain regions involved in the memory of cross-modal temporal order we compared the blood oxygenation level-dependent (BOLD) activation between the mixed and separate order conditions where number of items was same. At the same time, cortical areas sensitive to the memory load were mapped by comparing the high load with low load conditions in the separate recall task.

SUBJECTS AND METHODS

Subjects

Fourteen normal right-handed volunteers (five men; mean age, 21.3 years; age range, 20–22 years), with no history of psychiatric or neurologic disease, were recruited from Fudan Medical College. They participated in the experiment upon informed consent.

Materials and Tasks

There were three types of within-subject conditions: mixed order (M); separate order with low load (L); and separate order with high load (H). Each subject accomplished three scans, one for each condition. The order of these three conditions was balanced between subjects. The basic task in all three conditions was memorizing followed by immediate recall. The items to be memorized were a sequence of three or five digits (selected from 2–9) and three or five locations (out of nine possible locations). The digits were delivered by earphones and the spatial locations were presented with an LED panel, both controlled by a computer. The LED panel had nine red LEDs positioned at nine random locations. It displayed the spatial stimuli serially, one LED at a time. There was a small green LED just below each of the red LEDs. The green LEDs were on during the whole experiment, which marks the nine locations for subjects to point to. This design is similar to the one used in the Corsi block-tapping test [Lezak, 1986; p.453]. The LEDs were

distributed in an area of about 5 degrees (visual angle) across.

In a series of behavioral experiments, we replicated Penney's [1989] finding of the modality effect and showed that the performance of separate recall was the same (score or span was 12–13 items), whether the stimulus presentation order was mixed or separate. The performances of separate recall were much higher than that of the mixed recall (score: 6–7 items). These results suggest that the modality effect was due mainly to the recall order rather than to the stimulus presentation order [Zhang et al., 1997, 1999]. Stimulus order may therefore have a relatively small effect, if any, on fMRI measures. A better design would be one using mixed stimulus order in both mixed and separate recall tasks. It was used in our preliminary fMRI experiment, in which subjects were required to recall the items in either mixed or separate order according to a cue displayed at the beginning of each block of mixed or separate tasks. Two subjects made mistakes in the mixed task, however, and recalled items in separate order instead, because they failed to keep track of what recall task they should carry out inside the MR scanner. For practical considerations (i.e., not overloading the subject), in the present fMRI study, subjects were required to recall items in the same order as the presentation order, i.e., the recall order was defined by the stimulus presentation order. Specifically, auditory digits and visual locations were presented either with mixed order or with separate order. In the mixed condition, LED lights (L) and digits (D) were pseudorandomly interleaved (e.g., $D_1-L_1-L_2-D_2-L_3-D_3$). In the two separate conditions (L and H), visual locations were presented before digits (e.g., $L_1-L_2-L_3-D_1-D_2-D_3$). The stimulus list contained 6 items (3 lights and 3 digits) in both mixed and separate low-load conditions and 10 items (5 lights and 5 digits) in the separate high-load condition. Each stimulus, whether visual or auditory, lasted 300 msec. The presentation rate was 1 sec/item in separate low-load and mixed conditions, and 0.6 sec/item in the separate high-load condition, resulting in a 6-sec total presentation time for a trial in all conditions.

Subjects were required to remember the items and their presentation order, and to recall them immediately after the instruction cue (*Baogao*, a Chinese word meaning *report*; it instructed subjects to recall with the fingers). The cue was given 0.5-sec after the end of the last item. Subjects recalled the digits 2–9 with eight different finger gestures that were well known and familiar to all subjects. They recalled the locations of LEDs by pointing their index finger to the corresponding positions of the LEDs. The gestures for pointing and for signaling the digit 1 are quite similar; thus, we did not include this digit in the memory list. The recalls lasted about 4 sec in mixed and separate low-load tasks and 6 sec in the separate high-load task, and was followed by a long rest period to allow hemodynamic recovery.

Subject responses were recorded by one of the authors. It was not easy to accurately record subject's specific pointing positions. The index finger pointing indicated that the subject was responding to a given location (instead of a digit), so the

cross-modal order (i.e., the main factor in our study) could be recorded correctly. In a behavioral experiment carried out outside the MR scanner, subjects pressed keys below each red LED that were connected to a PC. In this case, subjects' responses were recorded accurately. The behavioral results are reported in the result section.

In the fMRI experiment, all subjects except two were able to accomplish the tasks while keeping their head still. In two subjects, their head motion was detected to be larger than 2 mm, so their data were excluded from further analysis. Data from 12 subjects thus were analyzed and reported in the following sections.

Data Acquisition

Images were obtained with a whole-body 1.5-T Siemens Magnetom Vision MR System (Erlangen, Germany) equipped with echo planar imaging (EPI) capability. A circularly polarized head coil was used, with padding added to restrict head motion. Functional images were acquired for each condition with an T2*-weighted EPI pulse sequence (TE = 51 msec, TR = 2 sec, field of view [FOV] = 24cm) with 16–17 axial slices (2 subjects: 17 slices with voxel size of $3.75 \times 3.75 \times 4.8$ mm; the other 10 subjects: 16 slices with the voxel size of $3.75 \times 3.75 \times 5$ mm) that covered the whole brain (occasionally small portions of ventral occipital areas were missed). Corresponding high resolution T1-weighted spin-echo (SE) (for anatomic overlay) and spoiled gradient-recalled-echo (SPGR) (for stereotaxic transformation, 155–180 sagittal slices; voxel size of $0.975 \times 0.975 \times 1$ mm) images were also collected.

One trial lasted 26 sec (13 TRs) including at least 14 sec of rest and a variable intertrial interval between 0–8 sec. The use of long rest and varying intervals was to remove potential interaction of adjacent trials [Dale, 1997], as well as to reduce the effect of habituation and anticipation [Liu et al., 2001; Rosen et al., 1998]. Each EPI scan included 16 trials and lasted 8 min 6 sec (243 images/slice). The first three images were discarded to account for the approach to steady state in the MR signal.

Data Analysis

The data were analyzed with AFNI (Analysis of Functional NeuroImage, online at <http://afni.nimh.nih.gov.afni>) [Cox, 1996]. First, the raw data was motion-corrected and normalized. Then the trials in the same condition were pooled together for generating the averaged mixed, separate high-load, and separate low-load epochs, respectively. The averaged epochs were 13-TRs long, trimming the variable intertrial intervals at the end.

A hemodynamic impulse response function published previously [Glover, 1999] was convolved with the three stimulus boxcar functions (Fig. 1). The convolved results were the modeled hemodynamic responses for the encoding and retrieval components of the three experimental conditions. These modeled responses (encoding and retrieval in each condition) were then used as regressors (independent variables) in multiple regression analysis. A partial *F*-test

Figure 1.

Three box-car stimulus functions and corresponding hemodynamic responses. The encoding component in mixed (M), separate low-load (L), and separate high-load (H) conditions had the same model responses (top row). The retrieval component in M and L had the same model responses (middle row). The retrieval component in H conditions was presumed to be longer (bottom row).

was used to determine the contribution of each component. Finally, activation maps were generated for each condition (M, H, and L, threshold; $P < 0.05$) and the two temporal components (encoding and retrieval), resulting in a total of six activation maps: three for encoding (M, L, and H) and three for retrieval (M, L, and H). A minimum cluster of four connected voxels was applied to these maps to reduce the isolated false activations. With the spatial clustering, the false positive level in this study was <0.0004 (AlphaSim.ps in AFNI).

For quantitative analysis, we focused on the dorsolateral PFC (Brodmann's area [BA]9/46), inferior frontal gyrus (FG) (BA44/45), premotor (BA6), superior and inferior parietal lobe (BA7 and BA40), superior temporal (BA22), insula, and superior colliculus, as these areas have been implicated in

processing of temporal information or cross-modal integration by some previous imaging studies. First, we identified these brain areas or regions-of-interest (ROI), making a mask that covered a given brain area (e.g., BA9 and BA46) based on the standard brain atlases [Talairach and Tournoux, 1988]. These standardized brain areas were then transformed into coordinates corresponding to each subject's fMRI scan. These individualized brain areas allowed us to quantitatively compare brain activations between subjects. Specifically, we calculated the activated volumes in each of these areas for the different experimental conditions.

To compare BOLD signal change across different conditions based on the same set of voxels, the three encoding maps (M, L, and H) were combined with a logical OR operator into a common encoding map. A common retrieval

Figure 2.

Average activated map of 12 subjects in mixed tasks in right PFC (BA9/46) and TPJ (BA40), left PFC (BA9/46), right BA6, and left BA7. The corresponding ROIs of these areas were bounded with the blank line.

TABLE I. Average accuracy and rates of order and item errors

Task	Accuracy (%)	Item error rate (%)		Order error rate (%)		
		Digit	Location	Within-modal		Cross-modal
				Digit	Location	
Mixed	80	0	0.8	0	0.8	18
Separate high-load	87	0	5	0	8	—
Separate low-load	99	0	0	0	0.8	—

map was derived in the same way. An average time course was extracted from voxels of the encoding and retrieval maps, respectively. Subjects encoded the stimuli during the first 6 sec, followed by retrieval (last 4 sec, in mixed and separate low-load, 6 sec in separate high-load condition). To account for hemodynamic delay of the BOLD response (4–6 sec), the average BOLD signal of frames 4–6 was defined as the encoding signal in all mixed and separate conditions, whereas the average BOLD signal of frame 7–8 was defined as the retrieval signal in mixed and separate low-load conditions and the average signals of frames 7–9 as the retrieval signal in the separate high-load condition. The average signal of frames 1, 2, 12, and 13 was taken as the baseline. This approach of data analysis is similar to that used by Chein and Fiez [2001] and Corbetta et al. [2002]. The activated volumes in those predefined ROIs were compared between the three conditions. If the activated volume in the mixed condition was significantly larger than that in the separate low-load condition in a given area, then the area was defined as sensitive to cross-modal order effect (order effect). If the volume of separate high-load was significantly larger than that of the separate low-load in a given area, then the area was defined as sensitive to load effect. In addition, the same comparison of BOLD signal change between the three conditions was carried out as well.

RESULTS

Behavioral Data

The average accuracy and error rates of 12 subjects in the behavioral experiment are displayed in Table I. The results indeed indicate that mixing temporal orders across modalities made the task of remembering temporal orders significantly more difficult. The mean accuracy of 12 subjects in mixed, separate low-load and separate high-load conditions were 80%, 99%, and 87%, respectively. Wilcoxon test show that the accuracy of mixed and separate high-load were significantly lower than was that of separate low-load ($z = 2.682$, $P = 0.007$; $z = 2.636$, $P = 0.008$ respectively). The performance of mixed and separate high-load conditions, however, was not significantly different ($z = 1.316$, $P = 0.188$), which is the desirable condition for the fMRI experiment, because we wanted the separate high-load condition to match the difficulty level of the mixed condition.

It is clear from Table I that most errors in the mixed task were cross-modal (18%) rather than within-modal (1.6%, P

< 0.009). In addition, within-modal errors were not different between the mixed and the separate low-load tasks (1.6% vs. 0.8%, $P = 0.317$). In other words, when only calculating the same types of errors (i.e., item and within-modal order errors) as in the separate task, the accuracy of the mixed task was 98.4%, which was not different from that in the separate low-load condition (99.2%). It provides additional evidence supporting that the random order of stimulus presentation has no effects on encoding and retrieval of the items and their within-modal orders in mixed task. Compared to the separate task, the mixed task has an additional component: the cross-modal-order. The results suggest that, in terms of Baddeley's model, the mixed presentation order does not affect operations of the two separate subsystems (phonological loop and visuospatial sketchpad), but put a high demand on integration of the two subsystems for the cross-modal order.

In the separate high-load task, the errors were only in spatial modal; the order error (8%) was not significantly different from the item error (5%, $P = 0.248$), and this spatial-order error of separate high-load was more than that of separate low-load and mixed ($P < 0.02$).

fMRI Data

Table II lists activated volumes in PFC (BA9/46) and other ROIs in mixed, separate low-load and separate high-load conditions and P -values of comparisons between the conditions in both the encoding and the retrieval phases. Briefly, during the encoding phase, the mixed condition activated larger volumes than did the separate low-load condition in some unilateral cortical areas whereas the separate high-load condition activated larger volumes in many areas bilaterally. At retrieval, however, only left BA7 showed a larger activation in mixed than in the separate low-load condition.

Activated Volume at Encoding

Results listed in Table II show that the activated volume in the mixed condition was significantly ($P < 0.01$, Wilcoxon Test) larger than that in the separated low-load condition in right BA 9/46, BA6, and left BA7. In other words, a significant order effect was found in these areas, implying that they are involved in cross-modal order processing. Except for right BA9/46, however, the areas that showed an order effect also showed a significant load effect ($P < 0.01$). In

TABLE II. Activated volume and *P*-value of comparison

Brodmann's area	Left					Right				
	M	L	H	<i>P</i>		M	L	H	<i>P</i>	
				M-L	H-L				M-L	H-L
Encoding										
9/46	1,587	813	1,526	—	0.003	1,490	578	1,210	0.008	—
6	3,538	2,438	4,998	—	0.002	3,513	1,561	3,815	0.002	0.002
7	7,160	4,780	8,771	0.009	0.002	6,698	4,878	7,505	—	0.002
Retrieval										
9/46	2,149	1,345	1,726	—	—	1,345	948	867	—	—
6	8,489	8,263	9,089	—	—	6,167	6,421	6,997	—	—
7	7,511	5,308	6,164	0.010	—	5,049	3,116	4,025	—	—

Volume measurements are given in mm³. Significant *P*-values (<0.01) of comparison between conditions (M-L, H-L) were listed (—, not significant).

M, mixed condition; L, separate low-load condition; H, separate high-load condition.

short, both order and load effects were found in right BA6 and left BA7, but only order effect was seen in right BA9/46. In addition, a significant load effect was found also in left BA 9/46, BA6, and right BA7. No significant difference in activated volume between the mixed and separate high-load condition was found in any ROIs.

The activation was found in the right PFC in 11 of 12 subjects in mixed task. Although all activated areas were in the same ROI (right BA9/46), the specific locations of the activation varied between the subjects. This variation in PFC activation is not rare in fMRI studies of working memory. Variabilities like this were also seen in studies by Prabhakaran et al. [2000] and Rypma and D'Esposito [1999].

As TPJ (BA40), superior temporal sulcus (STS), insula, and superior colliculus were found to be involved in the integration across modalities [Calvert et al., 2001 for review], the activated volumes in these areas were also analyzed and compared between the three conditions. A marginally significant order effect ($P = 0.026$), but no significant load effect ($P = 0.655$) was found in right TPJ (BA40). Neither significant order ($P > 0.456$) nor load effect ($P > 0.583$) was detected in STS (BA22). The mean activated volumes in the mixed condition in insula, superior colliculus, and inferior PFC were smaller than the cluster threshold (four voxels), i.e., no significant activation was found in these structures in the present experiment.

Activated Volume at Retrieval

As shown in Table II during the retrieval phase there was much less difference in the activated volume between different conditions. Only left BA7 had a significantly larger activation volume in the mixed than in the separate low-load condition (i.e., order effect). No load effect was found in any areas.

Results of fMRI Signal Change

We also compared fMRI percent signal change between mixed and separate conditions. The results based on signal

change were basically the same as those based on volume. Significant order and load effects were found at encoding, but not at retrieval. An important result is that signal change of right PFC in mixed condition was significantly higher than in both separate low-load and high-load conditions ($P = 0.002$ and 0.003 , respectively), and the signal changes were not significantly different between the two separate conditions ($P = 0.136$) (Table III). These results strengthen the conclusion that order effect but not load effect was found in right PFC. A significant order effect was also found at left PFC and right TPJ ($P < 0.01$) (Table III) and a significant load effect at left PFC, bilateral premotor (BA6), and superior parietal cortex (BA7) ($P < 0.01$). The signal change of right TPJ in mixed condition was significantly higher than that in high-load ($P = 0.005$), which was similar to that of right PFC.

DISCUSSION

Comparison Between Encoding and Retrieval Phases

As described above, activities were decomposed into encoding and the retrieval phases. Our discussion has focused

TABLE III. BOLD signal change and *P*-value of comparison

Brodmann's area	M (%)	L (%)	H (%)	<i>P</i>		
				M-L	H-L	M-H
9/46 Left	1.45	0.72	1.01	0.002	0.002	—
9/46 Right	1.42	0.84	0.99	0.002	—	0.003
40 Right	1.71	1.26	1.15	0.003	—	0.005

Significant *P*-values (<0.01) of comparison between conditions (M-L, H-L, M-H) were listed (—, not significant).

BOLD signal, blood oxygenation level-dependent signal; M, mixed condition; L, separate low-load condition; H, separate high-load condition.

mainly on encoding. At retrieval, however, the only significant effect was the order effect found in left BA7, and no load effect was found in any area. One possible explanation for this enhanced activation in left BA7 may be that it was related to right hand movement during retrieval. A similar finding was reported by Rowe and Passingham [2001]. Their fMRI study of working memory for location and time (in visual modality) revealed a left lateralized BA7 activation related to joystick usage during response. In any case, the results show a marked difference between encoding and retrieval. Although significant order and load effects were found in areas composing large-scale frontoparietal networks during encoding, there was almost no order or load effect during retrieval. Similar findings have been reported in other studies. For example, Klingberg and Roland [1998] observed right prefrontal activation during encoding but not during retrieval in a sound-picture paired-associates memory task, and Rypma and D'Esposito [1999] observed the load effects of working memory in dorsal PFC in only the encoding period.

Why was the order effect found mainly during encoding but not during retrieval? Penney [1989] pointed out that "...subjects will not easily remember the order of items present in different sensory modalities," and suggested that "presentation modality provides a strong basis for organization—the preferred or optimal organizational scheme. It is not merely a reflection of some discretionary strategy adopted by subjects, but rather an inherent property of memory system." It thus is natural that at encoding for the separate conditions, subjects adopted modality organization, the optimal strategy, which provided within-modal order. In contrast, it was necessary that subjects completed the hard integration for the cross-modal order, and they had to give up the optimal strategy in the mixed task. The encoding in mixed condition would thus be more difficult than that in the separate conditions.

The present imaging data show both order and load effects at encoding rather than retrieval, suggesting that the difficulty lies mainly at the encoding instead of the retrieval phase. This imaging finding is consistent with the results of a behavioral experiment, showing that divided attention had almost no effect on the retrieval but significant effect on encoding [Naveh-Benjamin et al., 2000].

Some neuroimaging experiments show that PFC activation is associated with response selection as well as error monitoring. For example, Rowe et al. [2000, 2001] found that the selection of one location based on its order was associated with a distinct frontoparietal network, including dorsolateral (dl) PFC (BA46). Subjects were required to recall the items and locations in their presentation order in the present study, so recall or retrieval may need the similar response selection. The left hemisphere dominant activation was found at retrieval ($P < 0.05$) in our study and in other selection studies [Desmond et al., 1998; Iacoboni et al., 1996; Rowe and Passingham, 2001; Rubia et al., 2001]. The fact that somewhat more activation was found in the mixed and separate high-load than in the separate low-load condition

in left PFC ($P < 0.17$, not significant) is consistent with this view. A possible account for the difference not reaching significance may be that the control task (separate low-load) in the present study required selection too, albeit with less demand. In contrast, the control task did not need selection at all in the studies of Rowe et al. [2000, 2001]. Error monitoring and detection is an executive function associated with performance or action monitoring [Gehring and Knight, 2000; Menon et al., 2001]. fMRI studies reported error-related brain activation in the bilateral [Menon et al., 2001] or left inferior frontal gyrus (IFG) [Garavan et al., 2002] and event-related potential (ERP) studies found that error-related negative (ERN) signal occurs at the moment of making an error [Gehring and Knight, 2000]. In our study, the order effect in PFC was founded only at encoding, not at retrieval (i.e., action or response), and right PFC (BA9/46) was more sensitive to the order effect. Error monitoring therefore is not likely the main cause for the activation difference at encoding.

Active Integration During Mixed Recall

The activation of the brain areas in the mixed condition was stronger than that in separate conditions at encoding but not at retrieval. This finding may help us to understand the role of active integration during mixed recall.

Baddeley [2000, 2003] proposed that, "We need to be able to separate the relatively automatic binding of properties that occur in the processes of normal perception from the more active and attentionally demanding integrative processes that are assumed to play such an important role in the episodic buffer," "the central executive, which is responsible for binding information from a number of sources into coherent episodes. Such episodes are assumed to be retrievable consciously ..." [Baddeley, 2000], and "...conscious awareness provides a convenient retrieval process." [Baddeley, 2003]. The present results can be interpreted based on Baddeley's new model, with the episodic buffer as follows. The central executive function is engaged in the mixed condition to coordinate (e.g., switch focus of attention between) the verbal and visuospatial subsystems for integrating the cross-modalities information at encoding. This is a more active and attentionally demanding integrative process. The product/output of the integration would be the episodes (either the cross-modal order or integrated temporal coding, which is similar to that of the episodic memory). [Tulving, 1983; p.38] and would be stored in the episodic buffer. Subsequently, these conveniently retrievable episodes would be used for retrieval in the mixed-recall condition, and this retrieval would not be very different from that of the separate conditions. In both cases, there would be no active integration during the retrieval phase. This view is consistent with the findings of this and other studies, that the right PFC is more activated during encoding but during not retrieval, in the mixed condition of the present study and in the sound-picture paired-associates memory task [Klingberg and Roland, 1998], and that the order-based working memory task effect (alphabetical order vs. forward order) was

TABLE IV. Areas showing the order effect at encoding

Brodmann's area	Talairach coordinates			Present study		Other studies	
	<i>x</i>	<i>y</i>	<i>z</i>	Order effect	Load effect	Cross modal	Temporal info
9/46 Right	35.4	33.0	28.2	Yvs	—	Y	Y
40 Right	50.3	−33.3	37.5	Ys	—	Y	Y
9/46 Left	35.5	33.8	26.9	Ys	Yvs	—	Y
6 Right	18.8	3.9	56.0	Yv	Yvs	—	Y
7 Left	−20.1	−55.1	49.7	Yv	Yvs	—	Y

Yv, Ys, and Yvs denote that Yes, a given effect was revealed from volume (Yv), signal (Ys), or both data (Yvs). The coordinates of activated areas at encoding in mixed condition are listed.

found in dl-PFC (mainly right side) during the delay/manipulation period, but not at retrieval [see Fig. 2 of Postle et al., 1999].

Conjunction of Temporal Order Processing and Cross-Modal Processing

Based on spatial volume and intensity of the BOLD responses, our data indicate that there was a significant order effect in bilateral PFC (BA9/46), right BA6, TPJ (BA40), and left BA7. These areas are potentially involved in cross-modal order processing. With the exception of the right PFC and TPJ, however, these areas also had larger volumes and high signal change of activation for the high load condition, exhibited a load effect. Combining the order and load effects, we divided these areas into two types: (1) area showing the order effect only (right PFC, TPJ; Fig. 2, top row); (2) area showing strong load effect in addition to the order effect (left PFC, right BA6, and left BA7; Fig. 2, bottom row).

We compared the present results to that of the four previous fMRI or positron emission tomography (PET) studies on temporal or order information processing within a modality [Chein and Fiez, 2001; Marshuetz et al., 2000; Rao et al., 2001; Rowe and Passingham, 2001], and four previous studies on cross-modal (visual-auditory) processing, which were not related to temporal order [Calvert et al., 2000; Downar et al., 2000; Klingberg and Roland, 1998]. Table IV shows the pattern that resulted from this comparison. Type 1 areas, the right PFC and TPJ, which showed the cross-modal order effect in our study, were involved in both the temporal information processing and the cross-modal tasks in the imaging studies cited above. In contrast, Type 2 areas, left PFC, right BA6, and left BA7, which showed both order and load effect in our study, were activated only in the temporal tasks in the cited studies.

Existing studies have shown that the right PFC and TPJ are involved in both cross-modal and temporal processing, which make it suitable for the memory of cross-modal orders, as demonstrated by the current study. Figure 2 shows the activated map of 12 subjects carrying out the mixed tasks, exhibiting significant activation in the right PFC and TPJ. Bushara et al. [2001] have reported a similar finding that the right PFC and inferior parietal lobule (BA40) were activated in a cross-modal temporal task (i.e., to detect the onset

asynchrony of auditory and visual stimuli). In contrast, left PFC, right BA6, and left BA7 were activated only in the temporal tasks. In the present study, left PFC, right BA6, and left BA7 showed both the order and the load effects. Considering together the pattern of results from the Bushara et al. [2000] study and the current study, it seems more likely that the enhanced activation in left PFC, BA7, and right BA6 in the cross-modal order task is due to the high demand on temporal information processing, rather than to the involvement of these areas in cross-modal processing. The analysis of error rates (Table I) shows that, compared to separate low-load, mixed and separate high-load tasks had significantly more errors of the cross-modal order and the within-modal order, respectively. Thus, both the mixed and separate high-load tasks had higher demand of temporal order processing than that in the separate low-load task.

Is the Right PFC the Neural Correlate of the Episodic Buffer?

In the mixed condition, subjects recalled auditory digits and visual locations with an intermixed order. In Baddeley's [2000] model, the episodic buffer best serves this cross-modal function. The results of the present study suggest that the right PFC may play a special role for this cross-modal integration function. Another recent study from Prabhakaran et al. [2000] also provided evidence supporting the role of the right PFC in episodic buffer. In the Prabhakaran et al. [2000] study, subjects were asked to maintain both spatial and verbal information either in an integrated or in an unintegrated fashion. In both conditions, subjects saw a target display of four letters and four spatial locations. In the integrated condition, the four letters to be remembered were displayed in the four locations to be remembered; thus, verbal information and spatial information were bound together. In the unintegrated condition, the four letters were presented centrally independent of the four locations; thus, verbal information and spatial information were separate. They found that only right PFC (right middle and superior frontal gyri) was involved in integrating information, and posterior regions were more active in separate tasks, which were more difficult.

The results of Prabhakaran et al. [2000] and the present study are similar in that both studies have identified in-

involvement of the right PFC in the processing of integrated information; and neither the current study nor the study by Prabhakaran et al. [2000] found significant load effect in the right PFC. It suggests that the right PFC seems to be more specific for this integration function.

The two studies were also different in some critical ways. The study by Prabhakaran et al. [2000] focused on the integration of information at the same spatial locations and presented both the letter and location information within the visual modality, making their integrated task easier than the unintegrated (separate) task. The current study focused on the integration of temporal order information across visuo-auditory modalities, with the mixed task harder than separate task. Despite these differences, both studies identified the right PFC as a key region for information integration. In terms of cognitive conjunction [Hirsch et al., 2001; Rubia et al., 2001], the right PFC may support a general function of working memory: information integration across sensory domains or across the verbal and spatial subsystems. Cross-domain integration, be it space-based or time-based, are functions of the episodic buffer. As Baddeley [2000] proposed, episodic buffer provides temporary storage of information held in a multimodal code, which is capable of binding information from the subsidiary systems. The episodic buffer has explicit functions in various integration tasks. Prabhakaran et al. [2000] have provided biological evidence for integration based on spatial location. The present study reveals neural correlates for the temporal integration across auditory and visual modalities. Together, the two studies complement each other and provide a more complete picture of the neural correlates of the episodic buffer.

Compared to the study by Prabhakaran et al. [2000], more areas were identified as supporting integration in the present study, including left PFC, right BA6, left BA7, and right TPJ. The reason may be that in contrast with integration in the previous study [Prabhakaran et al., 2000], the integration task here was composed of both cross-modal binding and memory of temporal order, and was more difficult than the separate task. As described above, left PFC, right BA6, left BA7, and right TPJ have been implicated in processing of temporal information, and right TPJ involved in cross-modality processing as well. These areas and right PFC thus constitute units of the large-scale frontoparietal network for memory of cross-modal temporal order.

We have compared the location (i.e., Talairach coordinates) of the right PFC activation in this study to that in other studies of executive functions [Bunge et al., 2000; Glahn et al., 2002; Postle et al., 1999; Veltman et al., 2003] and found good agreement. Given that the present result suggests right PFCs involvement in the episodic buffer, the agreement indicates that the episodic buffer and central executive (CE) might involve overlapping brain regions. This observation is analogous to other imaging findings that “storage of information occurs in the same neural ensembles that were involved in processing the original information” [Magnussen, 2000]. Our finding thus supports Baddeley’s

[2000, 2003] view that the episodic buffer is the inherent part of CE.

The signal change data show that right TPJ was similar to PFC (i.e., order but no load effect). Several imaging studies have shown that right PFC plays a critical role in executive functions but some experiments revealed widespread activation in frontal, parietal and temporal areas, leading Kübler et al. [2003] to suggest that executive functions may not be restricted to PFC; Baddeley [2000] holds a similar view for the episodic buffer. What the same and different functions are that the PFC and TPJ serve in the buffer and CE remains an interesting and open question.

The mixed task used in the current study certainly involved integration of the two-slave system of working memory, as the task required the cross-modal binding and integration of the verbal and spatial domains. Because integration of information across sensory modalities may not have the same neural mechanism as integration across different feature domains, it remains an important research question to differentiate further the contribution of cross-modal binding from that of integration of the two domains within a modality.

CONCLUSIONS

Contrasting between tasks requiring memory of cross-modal and within-modal orders, results from the present study show that the right and left prefrontal cortex (BA9/46), the right premotor (BA6), left parietal (BA7), and right TPJ (BA40) cortices had a stronger response in the mixed cross-modal task than in the separate task, indicating that these brain regions are involved in cross-modal temporal integration. These results and that of the study by Prabhakaran et al. [2000] complement each other in revealing the neural correlates of spatial- and temporal-based integration between the two subsidiary systems in working memory. Both studies identified the key role of right PFC in the integration, with the current study showing a more extensive network of areas involved in the cross-modal integration of temporal events, which is a more active and attentionally demanding integrative process [Baddeley, 2000].

ACKNOWLEDGMENTS

We thank the reviewers for their very helpful comments and suggestions on the early version of this article.

REFERENCES

- Baddeley A (1986): Working memory. Oxford: Oxford University Press.
- Baddeley A (1992): Working memory. *Science* 255:556–559.
- Baddeley A (2000): The episodic buffer: a new component of working memory? *Trends Cogn Sci* 4:417–423.
- Baddeley A (2003): Working memory: looking back and looking forward. *Nat Rev Neurosci* 4:829–839.
- Badgaiyan RD, Schacter DL, Alpert NM (2001): Priming within and across modalities: exploring the nature of rCBF increases and decreases. *Neuroimage* 13:272–282.

- Bunge SA, Klingberg T, Jacobsen RB, Gabrieli JDE (2000): A resource model of the neural basis of executive working memory. *Proc Natl Acad Sci USA* 97:3573–3578.
- Bushara KO, Grafman J, Hallett M (2001): Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci* 21:300–304.
- Bushara KO, Hanakawa T, Immisch I, Toma K, Kansaku K, Hallett M. (2003): Neural correlates of cross-modal binding. *Nat Neurosci* 6:190–195.
- Calvert GA, Campbell R, Brammer MJ (2000): Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649–657.
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001): Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14:427–438.
- Chen JM, Fiez JA (2001): Dissociation of verbal working memory system components using a delayed serial recall task. *Cereb Cortex* 11:1003–1014.
- Corbetta M, Kincade MJ, Shulman GL (2002): Neural systems for visual orienting and their relationships to spatial working memory. *J Cogn Neurosci* 14:508–523.
- Corbetta M, Shulman GL (2002): Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:215–229.
- Cox RW (1996): AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Dale AM, Bruckner RL (1997): Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp* 5:329–340.
- Desmond JE, Gabrieli JDE, Glover GH (1998): Dissociation of frontal and cerebellar activity in a cognitive task: evidence for a distinction between selection and search. *Neuroimage* 7:368–376.
- Downar J, Crawley AP, Mikulis DJ, Davis KD (2000): A multimodal cortical network for the detection of changes in the sensory environment. *Nat Neurosci* 3:277–283.
- Driver J, Spence C (2000): Multisensory perception: beyond modularity and convergence. *Curr Biol* 10:731–735.
- Garavan H, Ross TJ, Murphy K, Roche RAP, Stein EA (2002): Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *Neuroimage* 17:1820–1829.
- Gehring WJ, Knight RT (2000): Prefrontal-cingulate interactions in action monitoring. *Nat Neurosci* 3:516–520.
- Glahn DC, Kim J, Cohen MS, Poutanen V, Therman S, Bava S, Van Erp TGM, Manninen M., Huttunen M, Lönnqvist J, Standertskjöld-Nordenstam CG, Cannon TD (2002): Maintenance and manipulation in spatial working memory: dissociations in the prefrontal cortex. *Neuroimage* 17:201–213.
- Glover GH (1999): Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* 9:416–429.
- Hirsch J, Moreno DR, Kim KHS (2001): Interconnected large-scale system for three fundamental cognitive tasks revealed by function MRI. *J Cogn Neurosci* 13:389–405.
- Iacoboni M, Woods RP, Mazziotta JC (1996): Brain-behavior relationships: Evidence from practice effects in spatial stimulus-response compatibility. *J Neurophysiol* 76:321–331.
- Klingberg T, Roland PE (1998): Right prefrontal activation during encoding, but not during retrieval, in a non-verbal paired-associates task. *Cereb Cortex* 8:73–79.
- Kübler A, Murphy K, Kaufman J, Stein EA, Garavan H (2003): Co-ordination within and between verbal and visuospatial working memory: network modulation and anterior frontal recruitment. *Neuroimage* 20:1298–1308.
- Laurienti PJ, Wallace MT, Maldjian JA, Christina MS, Barry ES, Jonathan HB (2003): Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Hum Brain Mapp* 19:213–223.
- Lzak M (1983): *Neuropsychological assessment* (2nd ed.) New York: Oxford University Press. p 453–454.
- Liu TT, Franck LR, Wong EC, Buxton RB (2001): Detection power, estimation efficiency, and predictability in event-related fMRI. *Neuroimage* 13:759–773.
- Magnussen S (2000): Low-level memory processes in vision. *Trends Neurosci* 23:247–251.
- Marshuetz C, Smith EE, Jonides J, DeGutis J, Chenevert TL (2000): Order information in working memory: fMRI evidence for parietal and prefrontal mechanisms. *J Cogn Neurosci* 12(Suppl): 130–144.
- Menon V, Adelman NE, White CD, Glover GH, Reiss AL (2001): Error-related brain activation during a Go/NoGo response inhibition task. *Hum Brain Mapp* 12:131–143.
- Naveh-Benjamin, Craik FI, Perretta JG, Tonev ST (2000): The effects of divided attention on encoding and retrieval processes: the resiliency of retrieval processes. *Q J Exp Psychol A* 53:609–625.
- Penney C (1989): Modality effects and the structure of short-term verbal memory. *Mem Cognit* 17:398–422.
- Postle BR, Berger JS, D'Esposito M (1999): Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proc Natl Acad Sci USA* 96:12959–12964.
- Prabhakaran V, Narayanan K, Zhao Z, Gabrieli JD (2000): Integration of diverse information in working memory within the frontal lobe. *Nat Neurosci* 3:85–90.
- Rao SM, Mayer AR, Harrington DL (2001): The evolution of brain activation during temporal processing. *Nat Neurosci* 4:317–323.
- Rosen BR, Buckner RL, Dale AM (1998): Event-related functional MRI: past, present, and future. *Proc Natl Acad Sci USA* 95:773–780.
- Rowe JB, Josephs ITO, Frackowiak RSJ, Passingham RE (2000): The prefrontal cortex: Response selection and maintenance within working memory? *Science* 288:1656–1660.
- Rowe JB, Passingham RE (2001): Working memory for location and time: activity in prefrontal Area 46 relates to selection rather than maintenance in memory. *Neuroimage* 14:77–86.
- Rubia K, Russell T, Overmeyer S, Brammer MJ, Bullmore ET, Sharma T, Simmons A, Williams SCR, Giampietro V, Andrew CM, Taylor, E (2001): Mapping motor inhibition: Conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage* 13:250–261.
- Rypma B, D'Esposito M (1999): The roles of prefrontal brain regions in components of working memory: Effects of memory load and individual differences. *Proc Natl Acad Sci USA* 96:6558–6563.
- Talairach J, Tournoux P (1988): *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Tulving E (1983): *Elements of episodic memory*. Oxford: Oxford University Press. 38 p.
- Veltman DJ, Rombouts SARB, Dolan RJ (2003): Maintenance versus manipulation in verbal working memory revisited: an fMRI study. *Neuroimage* 18:247–256.
- Zhang DR, Jiang X, Tang XW (1997): Mixed-modality span of visuo-spatial stimuli and auditory digits. *Acta Psychol Sin* 29:234–239.
- Zhang DR, Tang XW, Chen XC, Xie H (1999): Possible account for the decrement of recall according to temporal presentation order in visuospatial and auditory dual memory task. *Acta Psychol Sin* 31:1–6.