MIND, BRAIN, AND EDUCATION

A Functional Magnetic Resonance Imaging Study of Foreign-Language Vocabulary Learning Enhanced by Phonological Rehearsal: The Role of the Right Cerebellum and Left Fusiform Gyrus

Kai Makita[1,2], Mika Yamazaki[3], Hiroki C. Tanabe[1,2,4], Takahiko Koike[2], Takanori Kochiyama[5], Hirokazu Yokokawa[6], Haruyo Yoshida[7], and Norihiro Sadato[1,2,8]

ABSTRACT— Psychological research suggests that foreign-language vocabulary acquisition recruits the phonological loop for verbal working memory. To depict the neural underpinnings and shed light on the process of foreign language learning, we conducted functional magnetic resonance imaging of Japanese participants without previous exposure to the Uzbek language using learning of novel Uzbek words. During encoding, spoken Uzbek words and corresponding visual objects were shown, and subjects either overtly repeated the words (phonological rehearsal) or overtly rehearsed numbers (phonological suppression). Phonological rehearsal improved the encoding performance. A learning-related decrease in rehearsal-specific activation was found in the left fusiform gyrus, right inferior temporal gyrus, and right cerebellum. Recollection of the phonologically rehearsed words activated the right cerebellum and left fusiform gyrus more prominently than recollection of the phonologically suppressed words in a performance-dependent manner. The phonological loop might provide the temporal and fragile registration of the articulatory pattern that is converted into a more durable form in the right cerebellum, which is in turn integrated with the object information in the fusiform gyrus.

[1]Department of Physiological Sciences, The Graduate University for Advanced Studies (Sokendai), Okazaki 444-8585, Japan
[2]National Institute for Physiological Sciences (NIPS), Okazaki 444-8585, Japan
[3]Child Development Research Center, University of Fukui, Fukui 910-1193, Japan
[4]Department of Social and Human Environment, Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan
[5]The Hakubi Project, Primate Research Institute, Kyoto University, Kyoto 606-8502, Japan
[6]School of Language and Communication, Kobe University, Kobe 657-8501, Japan
[7]Department of English Education, Osaka Kyoiku University, Kashiwara 582-8582, Japan
[8]Biomedical Imaging Research Center, University of Fukui, Fukui 910-1193, Japan

Address correspondence to Norihiro Sadato, Section of Cerebral Integration, Department of Cerebral Research, National Institute for Physiological Sciences, Myodaiji, Okazaki, Aichi 444-8585, Japan; e-mail: sadato@nips.ac.jp

With an increase in the migration of workers across the world, achieving high levels of proficiency in a foreign language has become economically and socially important (Service, 1992).

Learning new words is a crucial part of learning a foreign language (Service, 1992) and is closely related to working memory (Buchsbaum & D'Esposito, 2008; Gathercole & Baddeley, 1993; Papagno, Valentine, & Baddeley, 1991), which consists of a visuospatial short-term memory, a verbal short-term memory, a central executive, and an episodic buffer (Baddeley, 2003, 2012). Verbal short-term memory has two subcomponents: a phonological store that temporarily holds the phonological representation and an articulatory control process that rehearses the phonological representation within the phonological store (Baddeley, 1997). The rehearsal process refreshes the memory traces that decay within 1–2 seconds; thus, the whole system is also known as the phonological loop. The activity of this loop can be measured by the digit span during immediate serial recall or nonword repetition (Gathercole, Willis, Baddeley, & Emslie, 1994).

The effective capacity of the phonological loop is diminished when list items have long names rather than short names (the word-length effect), have names that are phonologically similar to one another (the phonological similarity effect), and when participants are required to engage in irrelevant articulation during presentation of the memory list (the articulatory suppression effect). The word-length and articulatory suppression effects are located in the rehearsal processes and the phonological similarity effect in the phonological store. These effects indicate that the representation of the phonological loop is subject to temporal degrading and is therefore fragile (Baddeley, Gathercole, & Papagno, 1998).

Behavioral studies have shown the functional significance of the phonological loop in foreign-vocabulary acquisition. A study of 9- to 10-year-old Finnish children learning English as a foreign language demonstrated a close association between nonword repetition ability and English-language ability 3 years later (Service, 1992). In adults, suppression of rehearsal by uttering an irrelevant sound disrupted paired-associate learning in a foreign language, but not in their native language (Baddeley, 2003; Papagno et al., 1991). Vallar and Baddeley (1984a, 1984b) described patient "P.V.," who had a pure phonological immediate-memory deficit but normal language production and normal comprehension of individual words and short sentences. P.V. learned native language pairs as rapidly as normal control subjects but failed to learn to associate a familiar word with an unfamiliar item from another language (Baddeley, Papagno, & Vallar, 1988). These data strongly suggest that the phonological loop plays a role in learning new foreign words. However, the precise nature of the relationship between phonological short-term memory and nonword repetition and/or vocabulary learning has remained unclear (Gupta & Tisdale, 2009).

Functional neuroimaging studies demonstrate that the phonological store is supported by left parietal regions and the subvocal rehearsal system is associated with Broca's area (Fiez et al., 1996; Gruber, 2001; Paulesu, Frith, & Frackowiak, 1993; Smith, Jonides, Marshuetz, & Koeppe, 1998). The role of the cerebellum in the phonological loop has attracted attention (Chein & Fiez, 2001; Chen & Desmond, 2005a, 2005b; Price, 2010). Activation in the right superior cerebellum is related to phonological rehearsal and the right inferior cerebellum is associated with the phonological store (Chen & Desmond, 2005a). Using functional magnetic resonance imaging (fMRI), Breitenstein et al. (2005) showed that the neural substrates of acquiring a novel lexicon included the left hippocampus and the left fusiform gyrus. However, this study did not make clear the role of the phonological loop in recruiting these areas during the task. Therefore, the neural substrates of the effect of the phonological loop on novel word learning remain unclear.

This study probed the neural substrates of the interaction of the phonological loop with word–object association learning of a foreign language. Behavioral evidence suggests that the phonological loop provides the temporal, fragile-but-precise registration of phonological sequences that are recoded in a durable form (Baddeley et al., 1998) during association with the corresponding object. As the strength of the association increases during encoding, the activity of the neural substrates of associative learning decreases (Breitenstein et al., 2005; Grill-Spector, Henson, & Martin, 2006). Therefore, we hypothesized that overt pronunciation of a word (involving phonological rehearsal) during the learning phase would lead to a learning-related reduction in activity over time, compared with producing overt, incompatible utterances (phonological suppression). Furthermore, the effect of the phonological loop during encoding should be evident during recollection, as retrieval of the learned word taps the long-term memory of its phonological structure (Baddeley et al., 1998), reactivating brain regions that were operational during the encoding (Nyberg, Habib, McIntosh, & Tulving, 2000). Therefore, we examined the recollection-related activation within the areas that were associated with encoding and anticipated that the recollection-related activation would predict the performance. Specifically, we expected involvement of the cerebellum and the left fusiform gyrus during both encoding and retrieval. Mastering vocabularies is a type of procedural learning because it includes the successful pronunciation of novel phonological sequences. The lateral cerebellum represents both motor learning (encoding) and the learned internal model (long-term memory) of a movement (Imamizu et al., 2000). The left fusiform gyrus is known to be involved in the binding and integration of multimodal stimuli (Breitenstein et al., 2005; Buchel, Price, & Friston, 1998; Murtha, Chertkow, Beauregard, & Evans, 1999; Price & Devlin, 2003). Breitenstein et al. found that the acquisition of a novel lexicon activated the left fusiform gyrus, which was associated with increasing vocabulary proficiency. Thus, we predicted that the cerebellum and the left fusiform gyrus would be active during both the encoding and the retrieval of novel foreign words, in the overt rehearsal condition.

We conducted event-related fMRI with healthy, native Japanese speaking volunteers who had never been exposed to the Uzbek language. During encoding runs, spoken Uzbek words and corresponding visual objects were shown, and subjects either overtly repeated the words (phonological rehearsal) or overtly rehearsed numbers (phonological suppression). During recollection runs, subjects were presented with either the visual object or the spoken Uzbek word, and were required to recollect (without speaking) and report via button press the corresponding word or object.

METHODS

Subjects

Twenty-four healthy volunteers (6 men and 18 women; mean age = 22.1 years; SD = 4.3) participated in this experiment. All subjects were native speakers of Japanese, educated beyond college level, and right-handed (Oldfield, 1971) with normal or corrected-to-normal visual acuity and normal hearing. The Japanese educational system includes the teaching of English as a foreign language, which starts at the age of 12 years (junior high school). Thus, all of the participants had studied English (the average number of years spent studying English was 9.7 [SD = 2.9 years]; English proficiency was not assessed). No subject had stayed abroad for longer than 1 year or previously been exposed to the Uzbek language. None of the subjects had a history of neurological or psychiatric illness. The protocol was approved by the ethical committee of the National Institute for Physiological Sciences, Okazaki, Japan. This study was conducted according to the Declaration of Helsinki (World Medical Organization, "Declaration of Helsinki," available from http://www.wma.net/en/30publications/10policies/b3/index.html, 2008). All subjects gave their written informed consent for participation.

Stimulus Preparation and Presentation

We selected 60 concrete Uzbek nouns comprising three categories of animals, fruits, and vegetables, and human-made objects (Table 1). For the auditory stimuli, 60 Uzbek words pronounced by a native male speaker were digitally recorded. For the visual stimuli, pictures representing the meaning of the 60 Uzbek words were prepared by digitally scanning line-drawings by Snodgrass and Vanderwart (1980) and Nishimoto, Miyawaki, Ueda, Une, and Takahashi (2005). The 60 line-drawings were duplicated and scrambled. A pilot study confirmed that it was not possible to assign meaning to the scrambled pictures. For the experiment, 60 word–picture pairs were generated, 30 of which contained meaningful pictures and 30 of which contained scrambled pictures. All stimuli were presented using Presentation software (Neurobehavioral Systems, Albany, CA, USA) on

Table 1
The 60 Uzbek Words Used in This Study (With English Translation)

Animals		Fruits and vegetables		Human-made tools	
Uzbek	English	Uzbek	English	Uzbek	English
Moshkhurma	Squirrel	Tarbuz	Watermelon	Pechkash	Screwdriver
Peshai	Cat	Kadu	Pumpkin	Almare	Dresser
Khuuk	Pig	Jugare	Corn	Pechagh	Kitchen knife
Tewa	Camel	Toot	Strawberry	Piala	Cup
Chishqan	Mouse	Samaroq	Mushroom	Sat	Clock
Qurbaqqa	Frog	Zarjama	Carrot	Yakhchal	Refrigerator
Fil	Elephant	Mumpali	Peanut	Korpacha	Bed
Eilan	Snake	Keila	Banana	Rawak	Shelf
Tamsah	Alligator	Piaz	Onion	Qashigh	Spoon
Tauugh	Chicken	Alugelas	Cherry	Ainak	Glasses
Gorakhar	Zebra	Ananas	Pineapple	Chakkush	Hammer
Aat	Horse	Nakhud	Pea	Darwaza	Door
Kuchii	Dog	Uzum	Grapes	Qalam	Pen
Einai	Cow	Alma	Apple	Jarep	Broom
Donqez	Bear	Torb	Radish	Meiz	Desk
Zarafa	Giraffe	Shaftalu	Peach	Chawki	Chair
Asa	Bat	Pamajan	Tomato	Guldan	Vase
Eishai	Donkey	Kachalu	Potato	Uttu	Iron
Shadi	Monkey	Qauon	Melon	Qaichi	Scissors
Kaftar	Turtle	Naak	Pear	Kelkin	Window

a personal computer. Using a liquid-crystal display projector, the visual stimuli were projected onto a half-transparent viewing screen located behind the head coil, which the subjects viewed through a mirror. The auditory stimuli were presented through MRI-compatible headphones. For each subject, the volume of the sound was adjusted to an appropriate level for task execution. Subjects' vocalizations during encoding runs were digitally recorded via an MRI-compatible microphone-recording system. Responses during the recollection phase were recorded using an optical button-box. Throughout the runs, the subjects were asked to focus on a small, black crosshair placed at the center of the screen.

Experimental Design and Procedures

Overall Design
The experiment had both encoding and recollection phases. The encoding phase consisted of eight event-related fMRI runs, with the encoding task alternating eight times with the performance test. The recollection phase (consisting of block design fMRI runs) followed the completion of the encoding phase.

MRI Acquisition
MRI images were acquired on an Allegra 3 Tesla MR imager (Siemens, Erlangen, Germany). For anatomical reference, T1-weighted high-resolution images were collected. For the encoding phase, a time-course series of 80 volumes was acquired using a T2*-weighted gradient-echo echo-planar

imaging (EPI) sequence. Each volume consisted of 34 slices (thickness = 4.0 mm) with a 0.6-mm gap to cover the entire cerebral and cerebellar cortices. Oblique scanning was used to exclude the eyeballs from the images. The field of view (FOV) was 192 mm and the in-plane matrix size was 64 × 64 pixels. The TR was 4,500 milliseconds with an FA of 87° and a TE of 30 milliseconds. We adopted a "sparse sampling" technique (Hall et al., 1999) in which the cluster-volume acquisition time was 2,000 milliseconds, followed by a 2,500-millisecond silent period. During the recollection phase, the protocol for the image acquisition was identical to the encoding runs, except that the duration of the silent period was set to 2,000 milliseconds.

Encoding Phase

Each encoding run had an event-related design, consisting of 4,500-millisecond trials (Figure 1) with four task conditions and one resting (control) condition. In each encoding trial, a white screen with a black crosshair was presented for 1,600 milliseconds, followed by the instruction cue for 400 milliseconds, which indicated whether subjects should use overt rehearsal or recite numbers while encoding the subsequent stimuli. If the instruction cue was a three-figure number, the subject had to rehearse overtly the number instead of the heard word when the response cue was presented. If the instruction cue was a nonsense picture (the scrambled digits of the three-figure number), the subject had to repeat overtly the upcoming aurally presented Uzbek word. At the onset of the silent period, both the Uzbek word and a corresponding picture (either a meaningful object or a meaningless scrambled image) were presented simultaneously. The picture was presented for 2,400 milliseconds, and Uzbek words were presented for 999 milliseconds on average (range = 836–1,158 milliseconds; SD = 84 milliseconds). The response was cued by a change in the color of the fixation crosshair. At 100 milliseconds before the end of the silent period, the screen returned to show a black crosshair on a white screen. This procedure generated the following conditions: audiovisual word–object association, with phonological rehearsal (Condition 1); audiovisual word–object association, with phonological suppression (Condition 2); no association, with phonological rehearsal (Condition 3); and no association, with phonological suppression (Condition 4). In addition, we included a baseline trial of 4,500 milliseconds during which the black crosshair on the white screen was presented without any auditory stimuli, which required no response. Each run consisted of all four conditions, with 15 trials of each condition with different word–picture pairs, and 15 baseline trials, yielding 75 trials in total. Thus, the 60 Uzbek words were each presented eight times per subject throughout the encoding runs. During repetition, the order of the Uzbek words was pseudorandomized within each condition. To

control word length, we equalized the average word length across four conditions. The average word lengths of the groups comprising four conditions (15 words per group) were 999 milliseconds (SD = 97 milliseconds), 999 milliseconds (SD = 84 milliseconds), 999 milliseconds (SD = 77 milliseconds), and 999 milliseconds (SD = 62 milliseconds), respectively. The distribution of the words in the four conditions was counterbalanced across subjects. Immediately after each encoding run, a performance test was conducted. In each trial, a four-framed picture (labeled 1 to 4) and a voice stimulus were presented simultaneously. The picture was shown for 3 seconds. Subjects were instructed to choose which frame was associated with the heard word via a button press. If the word did not correspond to any of the pictures, the subjects selected the blank picture frame (Number 4). No feedback was provided. In total, 60 trials per performance test were run.

Recollection Phase

For the recollection phase, we adopted a block design to enhance the task-related activity (Figure 2). The stimuli were 30 pairs of Uzbek words and objects previously used in the learning conditions (Conditions 1 and 2) of the encoding runs. The word recollection run consisted of six rest epochs (20 seconds) alternated with six task epochs (seven trials, 28 seconds in length), and ended with a rest epoch. During each task epoch, five object pictures were presented, one every 2 seconds, during the silent period. Three task blocks consisted of the 15 words presented in the preceding fMRI session with objects and overt rehearsal, and three task blocks involved the words encoded with articulatory suppression in Condition 2. Subjects were required to recollect covertly the corresponding paired Uzbek word (task trials). In each epoch, two additional trials were inserted after the word-presenting trial as follows. An Uzbek word was presented aurally without the visual presentation of the object. The subject had to decide whether it was the partner of the preceding visually presented object, indicating their response via button press (yes, index finger; no, middle finger; chance level = 50%). This was to confirm whether the subjects were attending to the task (catch trials). The 20-second rest epoch consisted of five null trials during which a black crosshair was presented on a white screen. Subjects were instructed to focus on the crosshair during the rest epoch. No feedback was given. The object recollection run followed a similar design, except that the Uzbek word was presented aurally, and the subjects were required to recollect the corresponding paired object (Figure 2).

Data Analysis
Performance

Subjects' learning over the course of eight runs was calculated as the percentage of correct responses. Subjects' performance was assessed by a two-way (learning condition [Conditions
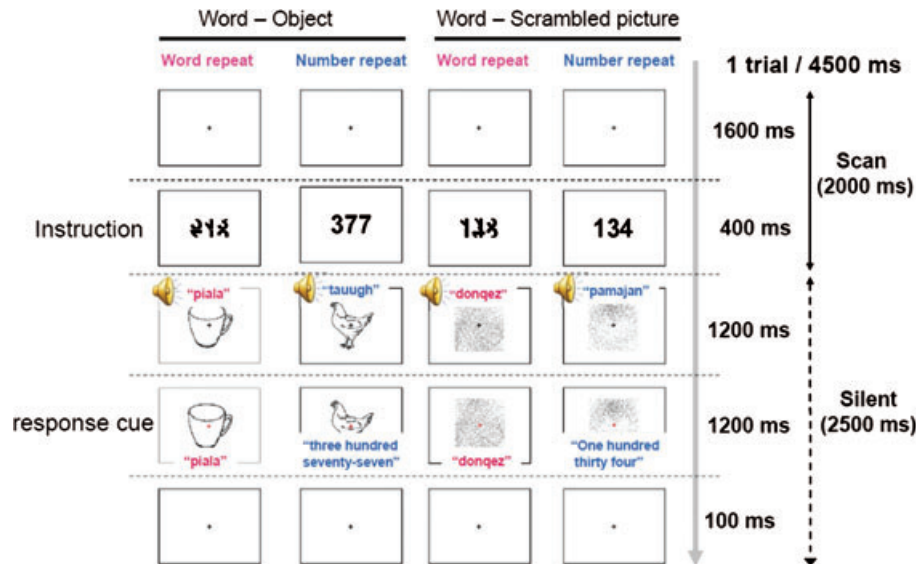
Fig. 1. Sample trials in different conditions of the encoding runs (event-related design). The time interval between two successive acquisitions of the fMRI images was 4,500 milliseconds. The volume acquisition time was 2,000 milliseconds (solid arrow), leaving a 2,500-millisecond silent period (dotted arrow). Before the audiovisual presentation of the word–object pair (left columns) or word–scrambled picture pair (right columns), an instruction cue was presented. When the cue was a scrambled picture, subjects had to overtly repeat the Uzbek word presented with either the meaningful object picture (far left, first column) or with a meaningless picture (third column). When the cue was a three-digit number, subjects had to overtly rehearse the number instead of the word (the second and fourth columns). The subjects' response was prompted by a change in color of the fixation cross from black to red (response cue), leaving 100 milliseconds to prepare for the next scan. In addition to these conditions, each of which included 15 words, there were 15 control trials without audiovisual presentation, yielding a total of 75 trials per run (337.5 seconds). Eight runs were completed.

1 and 2] and run [eight runs]) repeated measures analysis of variance (rmANOVA). To depict the learning effects clearly, the subjects were divided into two groups: those who achieved over 80% accuracy after association learning with phonological rehearsal were classified as "good learners" (4 males and 10 females) and the rest were classified as "poor learners" (2 males and 8 females).

fMRI Encoding Phase
The first two volumes from each fMRI encoding run were discarded, and the remaining 78 volumes per run were used for the analysis. The data were analyzed using Statistical Parametric Mapping software (SPM8; Wellcome Trust Centre for Neuroimaging, London, UK) implemented in MATLAB (Mathworks, Natick, MA, USA). The echo-planar images were realigned for motion correction, coregistered with the whole-head MP-RAGE image volume, which was then normalized to the Montréal Neurological Institute (MNI) stereotaxic space, and smoothed with an isotropic Gaussian kernel of 8-mm full-width-at-half-maximum in the x, y, and z axes. Statistical analysis was conducted at two levels. First, individual task-related activation was evaluated. Second, individual data were summarized and incorporated into a random-effects model. In the individual analyses, the signal time course for each subject was modeled for the conditions (1, 2, 3, 4, and baseline) by repetition (eight runs) using a delta function convolved with a hemodynamic response function, run effect, and high-pass filtering. To test hypotheses about regionally specific condition effects at each run, comparisons were made with the baseline condition of the same run using linear contrasts. The obtained contrast images of the condition by repetition effects (4 conditions $\times$ 8 repetitions $=$ 32 images per subject) were incorporated into a flexible factorial design that modeled the subject effect, the four different conditions, and the eight repetitions at the group level. As the learning effects were the main interest of the study, we focused on the good learners ($n = 14$) in this analysis of the learning-related changes in activation. To evaluate the learning-related changes in Conditions 1 and 2, we used the weighted contrasts calculated by the good learners' mean performance (accuracy) in each condition at each repetition. As we were interested in the effect of phonological rehearsal on learning, the contrasts were generated between the learning effects during Condition 1 and during Condition 2. Statistical significance was set at $p < .05$, corrected for multiple comparisons at the cluster level (Friston, Holmes, Poline, Price, & Frith, 1996). This procedure controlled family-wise Type 1 error strongly at the cluster level, permitting statistical inferences to be made about each cluster, and was based on the probability of getting a cluster

of the size observed (defined by a height threshold), or a larger one, in the volume analyzed. Cluster size was defined by the number of voxel which size was 2 mm × 2 mm × 2 mm.

Recollection Phase

The image preprocessing was similar to that for the encoding phase data. For the individual analyses, the signal time course for each subject was modeled with a boxcar function convolved with a hemodynamic-response function and high-pass filtering. Four explanatory variables were included in the model to test the effects of the recollection task (word/object) and the encoding condition of the recalled words (words encoded in Condition 1 [phonological rehearsal]/words encoded in Condition 2 [phonological suppression]). The catch trials were modeled as a regressor of no interest. Each task condition was compared with the rest condition of the same run to obtain contrast images. The individual contrast images of the good learners were entered into a 2 (Word Recollection/Object Recollection) × 2 (Words Learned With Rehearsal/Words Learned With Suppression) flexible factorial design matrix. The recollection-related activation was tested within the areas that showed the learning effect

specific to the phonological encoding condition. Finally, to evaluate the relationship between these regions and the vocabulary acquisition performance, for all subjects (including poor learners), correlation analyses were conducted between each subject's performance during the eighth test and their recollection-related activation.

RESULTS

Encoding Phase
Behavioral Performance

In terms of the subjects' behavioral performance ($n = 24$), three-way rmANOVA (Gender × Learning Condition × Run) revealed effects of learning condition, $F(1, 22) = 59.76$, $p < .001$, and run, $F(3.19, 70.19) = 55.92$, $p < .001$, and a significant interaction for Learning Condition × Run, $F(7, 154) = 3.52$, $p = .002$. No significant main effects of gender or its interaction were found (Gender × Learning Effect × Run, Gender × Learning Condition, Gender × Run). The good learners' ($n = 14$) mean accuracy at the end of the eighth run was 92.86% (8.04) for rehearsed words and 82.38% (14.93) for suppressed words. The poor learners' ($n = 10$) mean accuracy at the end
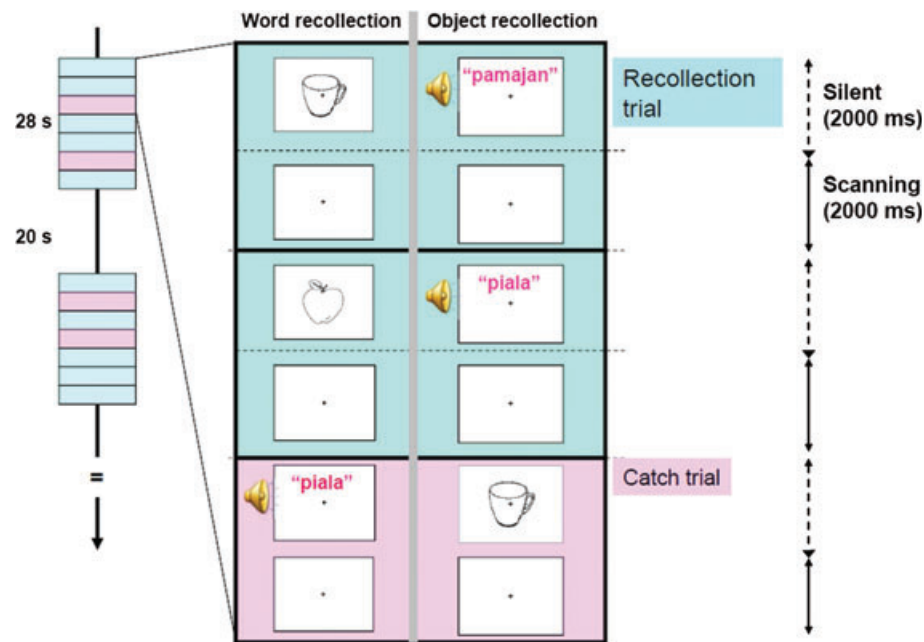

Fig. 2. Trial conditions of the recollection runs (mixed block and event-related design). Word recollection and object recollection were performed in separate runs. Each run consisted of three epochs using the pairs encoded during phonological rehearsal and three blocks using the pairs encoded during number rehearsal (phonological suppression). Each task epoch, 28 seconds in duration, was separated by a rest epoch consisting of 20 seconds of fixation. The time interval between two successive acquisitions of the fMRI images was 4,000 milliseconds. The volume acquisition time was 2,000 milliseconds (solid arrow), leaving a 2,000-millisecond silent period (dotted arrow) when a visual object (right column) or an auditory word (left column) was presented. The participants were asked to covertly recall the previously associated Uzbek word or object (task trial, blue rectangles) in five of seven trials during the task epoch. The remaining two trials were catch trials (pink rectangles), during which participants had to decide if the presented item was paired with the item presented in the previous trial using a yes/no button press (catch trial).
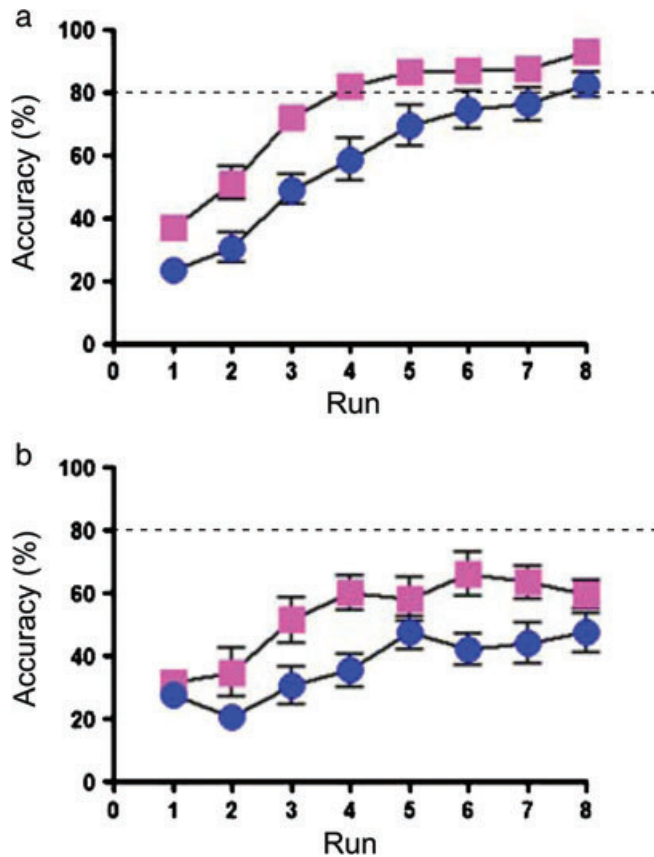
Fig. 3. Learning performance. Learning curves of the words encoded with phonological rehearsal (pink squares) and phonological suppression (blue circles) of (a) good learners ($n = 14$), whose performance reached over 80% at the eighth test for the words learned with rehearsal, and (b) the remaining subjects, classified as poor learners ($n = 10$). The x-axis indicates the tests conducted after each encoding run. The y-axis represents the mean accuracy score. Error bars represent the standard errors of the mean.

of the eighth run was 59.33% (14.21) for rehearsed words and 47.33% (20.23) for suppressed words (Figure 3). For the good learners, rmANOVA revealed significant effects of learning condition (rehearsed vs. suppressed), $F(1, 13) = 29.88$, $p < .001$, and run, $F(3.81, 49.47) = 92.39$, $p < .001$, and a significant interaction, $F(7, 91) = 2.41$, $p = .03$. For the poor learners, there was a significant effect of learning condition, $F(1, 9) = 67.69$, $p < .001$, and run, $F(7, 63) = 19.17$, $p < .001$, but no significant interaction, $F(7, 63) = 1.68$, n.s.

fMRI Results
Learning-Related Decrease in Activation
During word–object association learning with phonological rehearsal, good learners showed statistically significant signal decreases as learning progressed in the left inferior frontal gyrus, left medial frontal gyrus, left fusiform gyrus, right inferior temporal gyrus, and right cerebellum (Figure 4a).

To depict the regions with a rehearsal-specific learning effect, we compared the learning-related decline in activation during phonological rehearsal with that during phonological suppression within the areas that showed a rehearsal effect. The left fusiform gyrus, right inferior temporal gyrus, and cerebellum exhibited learning-related decreases specific to the phonological rehearsal condition (Figure 4b and 4c and Table 2).

Recollection Phase
As the encoding-related areas are involved in long-term memory formation, we examined the recollection-related blood oxygen level-dependent (BOLD) response of the left fusiform gyrus, the right inferior temporal gyrus, and the right cerebellum. During recollection of the words cued by the visual objects, the activation was more prominent for words learned during rehearsal than those encoded during suppression within the encoding-related areas (Figure 4d, left fusiform gyrus, $F(1, 13) = 5.54$, $p = .035$; right cerebellum, $F(1, 13) = 8.00$, $p = .014$; and right inferior temporal gyrus, $F(1, 13) = 5.05$, $p = .043$, rmANOVA). In these three regions, we also investigated the effect of performance on the BOLD response in all subjects ($n = 24$). For rehearsed words, subjects' performance was positively correlated with the recollection-related activation in the left fusiform gyrus (Pearson's $r = .44$, $p = .03$) and the right cerebellum ($r = .49$, $p = .02$). This correlation was not observed during retrieval of the phonologically suppressed words (Figure 4e).

DISCUSSION

As expected, phonological rehearsal improved association learning. A learning-related decrease in rehearsal-specific activation was found in the left fusiform gyrus, right inferior temporal gyrus, and right cerebellum. Recollection of the phonologically rehearsed words activated the right cerebellum and left fusiform gyrus more prominently than recollection of the phonologically suppressed words in a performance-dependent manner.

Improved Performance by Phonological Loop
In this study, the performance was improved with phonological rehearsal compared with phonological suppression. We interpreted this as the effect of verbal working memory on the cognitive task. The role of working memory in complex cognitive activities has been addressed by conducting experiments using the so-called dual-task interference paradigm (Baddeley, 2012; Shah & Miyake, 1999). In this paradigm, a cognitive task of interest is performed by itself and with a secondary task that is considered primarily to tap one of the subcomponents
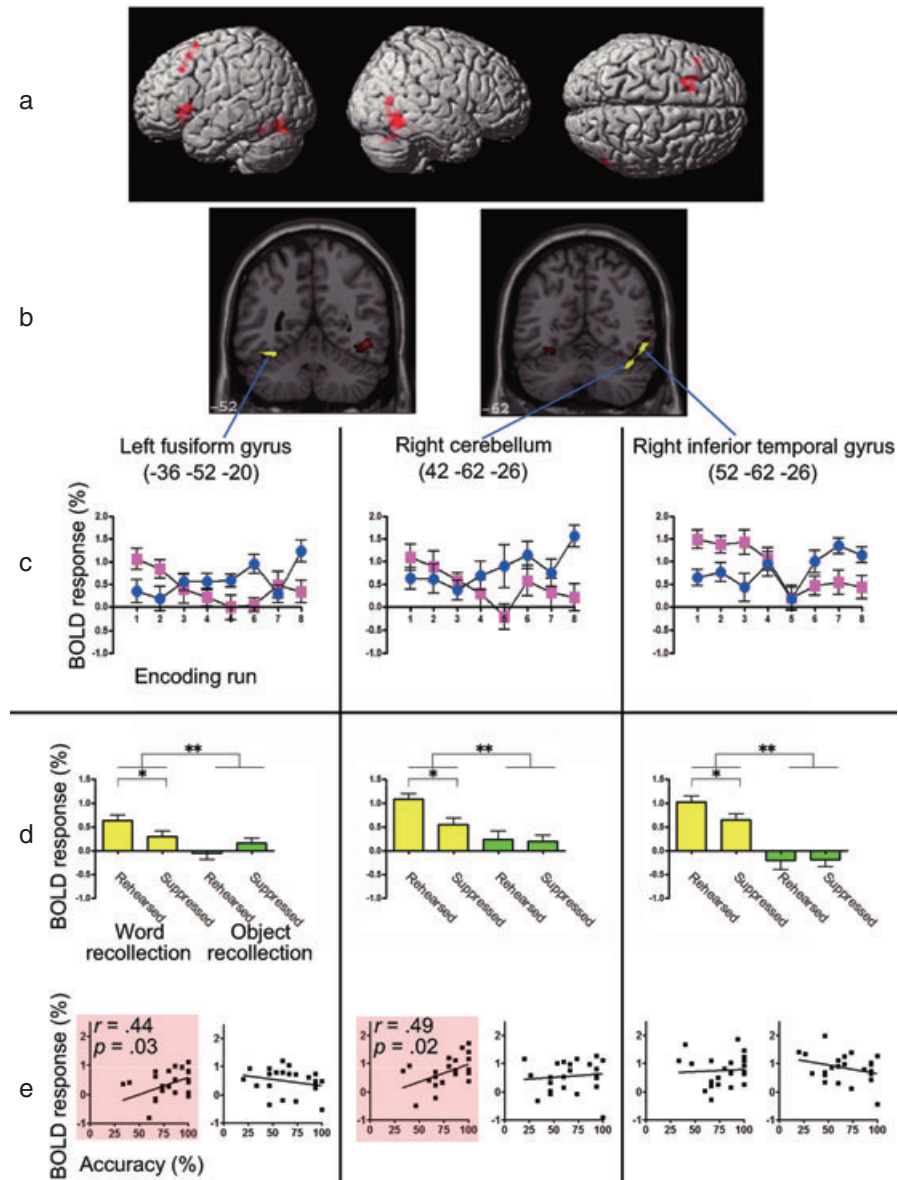
Fig. 4. Regions showing encoding- and recollection-related activation. (a) Regions exhibiting learning-related decline in activity during word–object association learning with phonological rehearsal (Condition 1), superimposed on the surface-rendered high-resolution MRI viewed from the left, right, and top. $p < .05$, corrected for multiple comparisons at the cluster level with whole-brain search. The cluster size corresponding to the corrected statistical threshold was 248 voxels within the analyzed volume of 179,633 voxels. (b) The regions showing rehearsal-specific learning-related decline in activity compared with Condition 2 (yellow) within the above-mentioned areas (red). The regions were superimposed on coronal sections of the T1-weighted high-resolution MRI normalized into MNI space at $y = -52$ mm (left) and $y = -62$ mm (right). $p < .05$, corrected for multiple comparisons at the cluster level with regional search. The cluster size corresponding to the corrected statistical threshold was 57 voxels within the analyzed volume of 2,109 voxels. (c) Learning-related decline of the BOLD response in the left fusiform gyrus (left), right cerebellum (middle), and right inferior temporal gyrus (right). The % BOLD response during encoding the word–object association with phonological rehearsal (pink squares) and with phonological suppression (blue triangles) for good learners ($n = 14$) is plotted. The error bars indicate the standard error of mean. (d) Recollection-related activation of these areas for good learners ($n = 14$). The yellow bars indicate the % BOLD response during word recollection, and the green bars show the % BOLD response during object recollection. The error bars indicate the standard error of the mean. $*p < .05$. $**p < .01$. (e) Correlation of the BOLD response during word recollection with encoding performance after the eighth encoding run. All subjects' data were included in the analysis ($n = 24$). In the left fusiform gyrus and the right cerebellum, performance was positively correlated with the BOLD response during the recollection of words that had been encoded during phonological rehearsal (highlighted in red). The x-axis indicates subjects' accuracy (%) during the eighth performance test, and the y-axis displays the BOLD response during the word recollection run.

Table 2
Brain Regions Showing Learning-Related Decline in Activation Specific to Encoding During the Phonological Rehearsal Condition

Cluster p-Value	Cluster Size	Voxel t-Value	MNI			Side	Location
			x	y	z		
Contrast: Learning-related decrease specific to phonological rehearsal (Condition 1 > Condition 2)							
0.03	90	4.44	−36	−52	−20	Left	Fusiform gyrus
0.01	165	4.19	52	−62	−12	Right	Inferior temporal gyrus
—	—	3.99	42	−62	−26	Right	Cerebellum lobule VIIa/Crus I

Note: The threshold was set at $p < .05$ corrected for multiple comparisons at the cluster level. The cluster size corresponding to the threshold was 57 voxels within the search volume of 2,109 voxels. Locations were defined using the SPM Anatomy Toolbox v1.8 (Diedrichsen, Balsters, Flavell, Cussans, & Ramnani, 2009; Eickhoff et al., 2005).

of working memory. If the secondary task disrupts the performance of the primary cognitive task when compared with the control condition, the subcomponent tapped by the secondary task is presumed to be involved in the performance of the primary cognitive task. This approach has been successfully used to specify whether a given cognitive task implicates a given subcomponent of working memory (Baddeley, 2012; Baddeley & Logie, 1999; Shah & Miyake, 1999).

In this study, the primary task was encoding word–object association, and the secondary tasks were rote rehearsal and number rehearsal. Both of the secondary tasks included the phonological loop process in the period between the instruction and the response cue (Figure 1); however, this was related to the heard words during rote rehearsal in contrast to the number during number rehearsal. Thus, differences in the performance of the primary task should reflect the involvement of the phonological loop in word–object association formation. Both conditions included the primary (making word–object association) and secondary (rote rehearsal vs. number rehearsal) tasks; however, rote rehearsal was closer to the primary task as it involved repetition of the heard word associated with the object. Thus, the effect of the phonological loop inevitably included the effect of attention that was part of the working memory, especially the central executive. However, it has been shown that articulatory suppression places minimal demands on executive processes but has a precise effect on the capacity for phonologically encoding the presented materials and for actively maintaining it by rehearsal (Baddeley, 2012; Baddeley et al., 1998). The present finding therefore suggests that the phonological loop supports the association learning of word–object pairs.

Task-Related Activation

The right cerebellum and the left fusiform gyrus showed a learning-related decrement in the activation during encoding that was specific to the phonological rehearsal condition. Furthermore, these areas showed retrieval-related activation that correlated with performance. This suggested the right cerebellum and the left fusiform gyrus as the areas where

the phonological loop interacts with long-term memory formation. In this study, learning-related activation was found in the lobule VIIa/Crus I in the posterior lobe of the right cerebellar hemisphere (Diedrichsen, Balsters, Flavell, Cussans, & Ramnani., 2009). This region is involved in articulatory control (Chen & Desmond, 2005a, 2005b) and motor learning. Imamizu et al. (2000) proposed that the motor learning-related activity in the cerebellum is explained by a computational theory in which the cerebellum acquires internal models of objects in the external world. The cerebellar activation increases significantly at the beginning of learning a new motor or cognitive task, and decreases as the learning proceeds, because the activation represents the error signals received by the multiple internal models. As learning proceeds, the error and cerebellar activation decrease. However, the task-related activation of the posterior lobe was present even after learning, when the error levels had been equalized. Imamizu et al. suggested that this cerebellar region reflects the acquired internal model of the motor skill. They also suggested that internal model formation in the cerebellum might occur for concepts, symbols, and languages. We extended this proposal to suggest that, if it was true, the internal model should be activated during recollection. Our data showed this to be the case. The right cerebellum was activated during the recollection of the aural Uzbek word when the associated visual object was presented. This cerebellar activation might represent the retrieved internal model of the pronunciation of the phonological sequences. This is supported by the finding that the effect of phonological rehearsal on retrieval was observed only when words, and not objects, were retrieved. As the cerebellum is known to support working memory by engaging inner speech mechanisms (i.e., the subvocal articulation of verbal content; Marvel & Desmond, 2010), the activation in the right cerebellum might represent the process of transforming the fragile temporal registration of the sound-pattern input from the phonological loop to the more permanent phonological structures.

The learning-related decline in the left fusiform gyrus during the encoding runs was specific to the phonological rehearsal condition. Hence, this decline cannot be explained simply

by the repeated visual presentations of objects but instead is related to the rehearsal-specific audiovisual association learning. Breitenstein et al. (2005) showed that the left fusiform gyrus is engaged during learning of aurally presented pseudowords and associated visually presented objects. As the association between the visual and phonological information was strengthened as a result of practice, the activity in the left fusiform gyrus decreased. The left fusiform gyrus has a role in the initial cross-modal integration of phonological, visual, and semantic information (Breitenstein et al., 2005; Buchel et al., 1998; Murtha et al., 1999; Price & Devlin, 2003). In blind listeners, enhanced speech-perception capabilities were associated with activation of the left fusiform gyrus and the primary visual cortex (Hertrich, Dietrich, Moos, Trouvain, & Ackermann, 2009). Involvement of the left fusiform gyrus in phonological processing is also supported by reports of patients presenting with phonological anomia accompanying lesions in this area (Foundas, Daniels, & Vasterling, 1998; Raymer et al., 1997). The left fusiform gyrus also contains orthographic representations of words (Binder, Medler, Westbury, Liebenthal, & Buchanan, 2006; Brunswick, McCrory, Price, Frith, & Frith, 1999; Buchel et al., 1998; Cohen et al., 2000, 2002). The present results suggest that the left fusiform gyrus is a cross-modal integration area that might also be related to visuomotor associations; that is, the integration of phonological and semantic information to generate supramodal "word-like" representations. The phonological loop might enhance the association of aurally presented words and visually presented objects through the correlated activation of the left fusiform gyrus and the right cerebellum, the latter of which is part of the phonological loop. To represent a multimodal event, several sensory regions must be involved, and must be interrelated. Nyberg et al. (2000) showed that recalling visual words that were paired with sounds at encoding activated some of the same auditory brain regions that had been engaged during the process. They suggested that a word–sound pair is a redintegrated multisensory whole. Here, redintegration refers to the relationship between any constituent parts of a complex whole and the totality of the whole. Redintegration is the classical psychological idea that is closely related to the concept of association. However, where association refers to the relation between parts that together form a whole, redintegration refers to the relation between any one of the constituent parts of a complex whole and the totality of the whole (Tulving & Madigan, 1970). Redintegrative recall was defined as a high conditional probability of recall of a "whole unit" given that a part of the unit has been recalled. In this study, an audiovisual and motoric "trio" could be seen as a multimodal whole that includes the representation of sound, vision, and the phonological articulation. This is redintegrated at retrieval by the unisensory (visual) presentation of the object. Therefore, it is conceivable that phonological rehearsal

enhances the memory of a foreign word as a whole through the addition of a motor component.

Limitations of the Study

There was a gender difference in this study. The difficulty in recruiting male participants resulted in an unbalanced gender distribution. As there was no gender effect on behavioral performance, we conducted imaging analysis without considering gender effects. Gender differences in the neural processing of language have been reported (Burman, Bitan, & Booth, 2008), and so this is to be investigated in future studies.

All of the participants in this study were native speakers of Japanese with experience of learning English as a second language (L2) within the Japanese educational system for 9.7 years on average. Therefore, they were not purely monolingual. As the proficiency level of English was not assessed, the effect of the proficiency of L2 (English) on the learning performance of L3 (Uzbek) could not be evaluated. This is another issue for future investigation.

CONCLUSION

The present results suggest that overt rehearsal enhances word memory by adding a motoric component to the multimodal whole and that this process is represented by the correlated activation of the right cerebellum and the left fusiform gyrus. This finding confirms the original notion by Baddeley et al. (1998) that the verbal working memory is essential for learning new foreign words. Furthermore, the neural substrates of the learning process revealed by this study suggest that vocabulary learning could be seen as obtaining audiovisual and motoric redintegration that can be retrieved in response to the cue of any one of the modalities. Finally, our findings illustrate the educational importance of rote rehearsal, based on the significance of vocabulary acquisition in learning a foreign language.

REFERENCES

Baddeley, A. D. (1997). *Human memory: Theory and practice* (Revised ed.). Hove, England: Psychology Press.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839. doi:10.1038/nrn1201.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. doi:10.1146/annurev-psych-120710-100422

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173. doi:10.1037//0033-295X.105.1.158

Baddeley, A. D., & Logie, R. H. (1999). The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge, England: Cambridge University Press.

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, *27*, 586–595. doi:10.1016/0749-596X(88)90028-9

Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *NeuroImage*, *33*, 739–748. doi:S1053-8119(06)00707-5

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*, *25*, 958–968. doi:10.1016/j.neuroimage.2004.12.019

Brunswick, N., McCrory, E., Price, C. J., Frith, C. D., & Frith, U. (1999). Explicit and implicit processing of words and pseudowords by adult developmental dyslexics: A search for Wernicke's Wortschatz? *Brain*, *122*, 1901–1917. doi:10.1093/brain/122.10.1901

Buchel, C., Price, C., & Friston, K. (1998). A multimodal language region in the ventral visual pathway. *Nature*, *394*(6690), 274–277. doi:10.1038/28389

Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience*, *20*, 762–778. doi:10.1162/jocn.2008.20501

Burman, D. D., Bitan, T., & Booth, J. R. (2008). Sex differences in neural processing of language among children. *Neuropsychologia*, *46*, 1349–1362. doi:10.1016/j.neuropsychologia.2007.12.021

Chein, J. M., & Fiez, J. A. (2001). Dissociation of verbal working memory system components using a delayed serial recall task. *Cerebral Cortex*, *11*, 1003–1014. doi:10.1093/cercor/11.11.1003

Chen, S. H., & Desmond, J. E. (2005a). Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. *NeuroImage*, *24*, 332–338. doi:10.1016/j.neuroimage.2004.08.032

Chen, S. H., & Desmond, J. E. (2005b). Temporal dynamics of cerebro-cerebellar network recruitment during a cognitive task. *Neuropsychologia*, *43*, 1227–1237. doi:10.1016/j.neuropsychologia.2004.12.015

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123*, 291–307. doi:10.1093/brain/123.2.291

Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain*, *125*, 1054–1069. doi:10.1093/brain/awf094

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage*, *46*, 39–46. doi:10.1016/j.neuroimage.2009.01.045

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*, 1325–1335. doi:10.1016/j.neuroimage.2004.12.034

Fiez, J. A., Raife, E. A., Balota, D. A., Schwarz, J. P., Raichle, M. E., & Petersen, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *Journal of Neuroscience*, *16*, 808–822.

Foundas, A. L., Daniels, S. K., & Vasterling, J. J. (1998). Anomia: Case studies with lesion localization. *Neurocase*, *4*(1), 35–43. doi:10.1080/13554799808410605

Friston, K. J., Holmes, A., Poline, J. B., Price, C. J., & Frith, C. D. (1996). Detecting activations in PET and fMRI: Levels of inference and power. *NeuroImage*, *4*, 223–235. doi:10.1006/nimg.1996.0074

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. New York, NY: Psychology Press.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*, *2*(2), 103–127. doi:10.1080/09658219408258940

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Science*, *10*(1), 14–23. doi:10.1016/j.tics.2005.11.006

Gruber, O. (2001). Effects of domain-specific interference on brain activation associated with verbal working memory task performance. *Cerebral Cortex*, *11*, 1047–1055. doi:10.1093/cercor/11.11.1047

Gupta, P., & Tisdale, J. (2009). Word learning, phonological short-term memory, phonotactic probability and long-term memory: Towards an integrated framework. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1536), 3755–3771. doi:10.1098/rstb.2009.0132

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., ... Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*(3), 213–223. doi:10.1002/(SICI)1097-0193(1999)7:3<213::AID-HBM5>3.0.CO;2-N

Hertrich, I., Dietrich, S., Moos, A., Trouvain, J., & Ackermann, H. (2009). Enhanced speech perception capabilities in a blind listener are associated with activation of fusiform gyrus and primary visual cortex. *Neurocase*, *15*(2), 163–170. doi:10.1080/13554790802709054

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Putz, B., ... Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature*, *403*(6766), 192–195. doi:10.1038/35003194

Marvel, C. L., & Desmond, J. E. (2010). Functional topography of the cerebellum in verbal working memory. *Neuropsychology Review*, *20*, 271–279. doi:10.1007/s11065-010-9137-7

Murtha, S., Chertkow, H., Beauregard, M., & Evans, A. (1999). The neural substrate of picture naming. *Journal of Cognitive Neuroscience*, *11*, 399–423. doi:10.1162/089892999563508

Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. *Behavior Research Methods*, *37*, 398–416. doi:10.3758/BF03192709

Nyberg, L., Habib, R., McIntosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Science USA*, *97*, 11120–11124. doi:10.1073/pnas.97.20.11120

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. doi:10.1016/0028-3932(71)90067-4

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30*, 331–347. doi:10.1016/0749-596X(91)90040-Q

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature*, *362*(6418), 342–345. doi:10.1038/362342a0

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*, 62–88. doi:10.1111/j.1749-6632.2010.05444.x

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, *19*, 473–481. doi:10.1016/S1053-8119(03)00084-3

Raymer, A. M., Foundas, A. L., Maher, L. M., Greenwald, M. L., Morris, M., Rothi, L. J., & Heilman, K. M. (1997). Cognitive neuropsychological analysis and neuroanatomic correlates in a case of acute anomia. *Brain and Language*, *58*(1), 137–156. doi:10.1006/brln.1997.1786

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology Section A*, *45*(1), 21–50. doi:10.1080/14640749208401314

Shah, P., & Miyake, A. (1999). Models of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory1* (pp. 1–27). Cambridge, England: Cambridge University Press.

Smith, E. E., Jonides, J., Marshuetz, C., & Koeppe, R. A. (1998). Components of verbal working memory: Evidence from neuroimaging. *Proceedings of the National Academy of Science USA*, *95*, 876–882. doi:10.1073/pnas.95.3.876

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. doi:10.1037//0278-7393.6.2.174

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Reviews of Psychology*, *21*, 437–484. doi:10.1146/annurev.ps.21.020170.002253

Vallar, G., & Baddeley, A. D. (1984a). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 151–161. doi:10.1016/S0022-5371(84)90104-X

Vallar, G., & Baddeley, A. D. (1984b). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive Neuropsychology*, *1*(2), 121–141. doi:10.1080/02643298408252018