From gestures to words: Spontaneous verbal labeling of complex sequential hand movements reduces fMRI activation of the imitation-related regions

Motohide Miyahara [a,b], Ryo Kitada [a,c,*], Akihiro T. Sasaki [a,c], Yuko Okamoto [a,c], Hiroki C. Tanabe [a,c,d], Norihiro Sadato [a,c]

[a] Division of Cerebral Integration, National Institute for Physiological Sciences, Okazaki, Japan
[b] School of Physical Education, University of Otago, Dunedin, New Zealand
[c] Department of Physiological Sciences, The Graduate University for Advanced Studies (Sokendai), Okazaki, Aichi, Japan
[d] Department of Social and Human Environment, Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan

ABSTRACT

The association of verbal labels with visuo-spatial patterns and sequences detectably alters neuronal activity in the brain in ways that have yet to be fully characterized. This study investigated the neural substrates involved in the effect of spontaneous verbal labeling on memorizing increasingly complex sequences of hand movements. We used functional magnetic resonance imaging (fMRI) to test our hypothesis that when verbal labels were employed, neuronal activity in imitation-related regions, such as the left inferior frontal gyrus (IFG), would be reduced, whereas without verbal labels, neuronal activation would increase. Sixteen healthy adults satisfactorily performed an immediate imitation task involving six levels of increasing complexity. After the fMRI experiment, participants reported at which complexity level they had formed verbal labels. Based on the self-report, we categorized the task blocks at each complexity level as either with verbal labeling (VL+) or without (VL−). Compared with VL+, the VL− condition activated the left IFG, bilateral middle frontal gyri, left precentral gyrus, and the right angular gyrus, whereas the opposite contrast revealed no significant brain activation. Verbal labeling seems to serve as an efficient heuristic that reduces the cost of cortical activation in the imitation-related regions.

© 2013 Elsevier Ireland Ltd and the Japan Neuroscience Society. All rights reserved.

1. Introduction

Sequence learning is a key element of voluntary behavior (Hikosaka et al., 1999) and a basis for higher cognitive functions, such as language (Rizzolatti and Arbib, 1998). To observe and imitate a long sequence of visuospatial items, such as a dance routine, we encode each sequential component in working memory as it occurs until the entire sequence is eventually memorized. In this process, simultaneous imitation of what is being observed is easier than delayed imitation (Caspers et al., 2010). In order to perform the latter — an exceptional human capacity of imitating a model action after a delay (Meltzoff, 1988) — we must wait until the model action is complete, use the visuospatial working memory to carry the perceived action forward, and reconstruct the model action; the more complex the sequence and the longer the delay, the more syntax and timing we must keep in mind and the more demanding the task (Barkley, 2004). The challenge of imitating complex

sequences can be met by using another uniquely human attribute, the ability for symbolic naming (Geschwind, 1964).

Our memory capacity can be expanded by naming visuospatial items; that is, encoding visual information into words. By recruiting verbal labels, the capacity of our visual working memory can be stretched from three, four (Owen, 2004), or five to seven digits (Baddeley et al., 1984). The effects of verbal labeling (VL) have been inferred from the reduced memory span of sequential hand movements that resulted from articulatory suppression (Frencham et al., 2003, 2004), and by the increased memory span that occurred after verbal labels were recruited for hand movements (Miyahara, 2007; Miyahara et al., 2008). These two methods provided evidence that VL contributes to the expansion of memory span for visuospatial items, such as sequential hand movements.

Before forming verbal labels for an immediate or a delayed imitation, the imitator needs to identify the elements, order, and length of the sequence. After recognizing the patterns of motor sequences or the structural template (Fiebach and Schubotz, 2006), the imitator might covertly verbalize the recognizable parts of the sequential movements (Nishitani et al., 2005). By labeling a movement element, the visual working memory (Haxby et al., 2000; Takahama et al., 2010) is assumed to be converted into verbal working memory.

* Corresponding author at: School of Physical Education, University of Otago, Dunedin 9054, New Zealand. Tel.: +64 3 479 8387; fax: +64 3 479 8309.
E-mail addresses: motohide.miyahara@otago.ac.nz (M. Miyahara), kitada@nips.ac.jp (R. Kitada).

Whether the modality of working memory is visual or verbal, an immediate imitation and a delayed imitation require on-line processing of the syntactic sequence (Fiebach and Schubotz, 2006); that is, a rule-based order of elements that can range from relatively simple (e.g., AA) to complex (e.g., ABCB) patterns. Functional magnetic resonance imaging (fMRI) studies have demonstrated greater activation in Broca's area during the processing of complex sequences compared with simple sequences for verbal sentences (Fiebach et al., 2005), and non-verbal sequences (Bahlmann et al., 2009). The activation of Broca's area seems to depend on the complexity of the syntactic sequence for both verbal and non-verbal sequences (Tettamanti et al., 2009). Broca's area has been described elsewhere as a flexible, general sequence and syntactic processor (Fiebach and Schubotz, 2006).

Complex non-verbal sequences can be remembered more easily by verbally labeling each sequential component, and then recalling the sequential labels verbally. This facilitation is analogous to automatization in learning a motor task which consists of a series of components. Each component may initially require substantial time and effort to perform, but all components are assimilated as a unified task whose individual components no longer require the same amount of time and effort as before (Schmidt and Lee, 1999). Although the involvement of verbal mediation has not been examined, previous fMRI research has demonstrated increased brain activation in the motor network for complex finger sequences (Sadato et al., 1996; Wu et al., 2004), and a reduction in activation once these movements have been well learned and have become automatic (Wu et al., 2004). After putting a verbal label on each automatized movement element, the imitator can execute the motor chunk associated with the label, thereby bypassing elaborate memory retrieval (Keele, 1973). VL of each motor chunk can thus be considered as partial automaticity, analogous to verbal mediation during the intermediate phase (Fitts, 1964), before full automaticity of the entire sequence is achieved during the late phase of learning. As is the case for fully automatized sequential finger movements (Wu et al., 2004), we reason that VL or partial automatization might reduce the brain activity cost associated with imitating sequential movements.

After verbally labeling each sequential component, the imitator still needs to memorize the entire sequence of labels verbally, especially in cases of long, complex sequences. Broca's area might have a role not only in the context of motor speech, which involves the hierarchical sequencing of linguistic syntax via verbal working memory (Fiebach et al., 2005; Santia and Grodzinskya, 2007), but also in the context of visual working memory (Haxby et al., 2000; Takahama et al., 2010) and sequence generation (Berkowitz and Ansari, 2008). Here, we regard Broca's area as a key locus for a domain-general sequence generator, which has the capacity to arrange all motor sequences, including those for speech (Fiebach and Schubotz, 2006). Among the motor sequences, we assume that articulatory sequencing is easier to perform, requiring a lower workload than the sequencing of movements of other body parts, such as the hand, because sequential order is more readily associated with verbal than spatial representations in working memory (Gmeindl et al., 2011). VL can consolidate sequential memory by associating a particular sequential component with a name and an articulation via the phonological loop. To generate a sequence for hand movements, for example, the workload of Broca's area should increase; VL of these hand movements might reduce the increasing workload. We tested the hypothesis that the activity of brain regions involved in the immediate imitation of hand movements would be reduced when participants employed verbal labels compared with when they did not use verbal labels for hand movements.

2. Methods

2.1. Participants

A total of 18 Japanese university students (four females and 14 males; aged 19–27 years) gave their written informed consent to take part in this study, which was approved by the Ethical Committee of the National Institute for Physiological Sciences of Japan. None of the participants reported any history of neuropsychiatric disorders or other contraindications for magnetic resonance imaging (MRI). All participants were right handed according to the Edinburgh handedness inventory (Oldfield, 1971).

2.2. fMRI data acquisition

A 3-T MR scanner (Allegra, Siemens, Erlangen, Germany) was used to acquire both T1-weighted magnetization-prepared rapid gradient-echo (MPRAGE) images (repetition time [TR] = 2000 ms, echo time [TE] = 4.38 ms, flip angle [FA] = 8°) and T2*-weighted gradient-echo echo-planar images (EPI; TR = 2000 ms, TE = 30 ms, FA = 75°, voxel size = 3 mm × 3 mm × 3.5 mm). A T1-weighted high-resolution anatomical image volume was obtained from each participant (voxel size = 0.898 mm × 0.898 mm × 1 mm) either before or after the run for acquiring the functional data. Each volume consisted of 34 slices of 3.5-mm thickness with 17% gaps, and covered the entire cerebral and cerebellar cortices.

2.3. Stimuli

The Kaufman Hand Movement Test (KHMT) (Kaufman and Kaufman, 1983) is used as a model of the present experimental task to observe and reproduce, in part because, as in our series, the test items in the KHMT are organized approximately in the order of increasing movement complexity in terms of span (i.e., numbers of hand gestures) and variation (i.e., types of hand gestures). The KHMT is an extension of Luria's apraxia test (Luria, 1966) which comprises three hand positions: side, fist, and palm (Fig. 1A). The KHMT starts with a sequence of two hand positions or a two movement span (i.e., side-side), and increases the number of hand positions (see Fig. 1B for an example). This order induces participants to form verbal labels spontaneously for each hand position or hand movement, as the visuospatial memory reaches its capacity limit, in order to cope with the memory demand by recruiting the phonological loop (Frencham et al., 2003, 2004; Miyahara, 2007; Miyahara et al., 2008).

Modeling after the KHMT, we developed the Okazaki Hand Movement Test (OHMT, Table S2) for the present fMRI experiment. The original 21 items of the KHMT were reorganized in order of increasing complexity (i.e., span and variation), to allow us to estimate the complexity level at which the verbal labels were formed. This required each complexity level to have the same span and variation. We initially generated two variations of each item from the KHMT by systematic replacements among the three hand positions. We then limited the range of serial hand positions to include only two, three, and four, instead of the range from two to six in the KHMT. Because the visual short-term memory capacity has been suggested to be three or four spans (Owen, 2004), we considered sequences longer than four to be unnecessary. Finally, we ordered the items such that no two consecutive items started with the same hand movement to avoid the participant's sense of immediate déjà vu which could have contaminated the brain activation of interest. Consequently, the OHMT consisted of 60 items for a single fMRI run (Fig. 1B and Table 1).

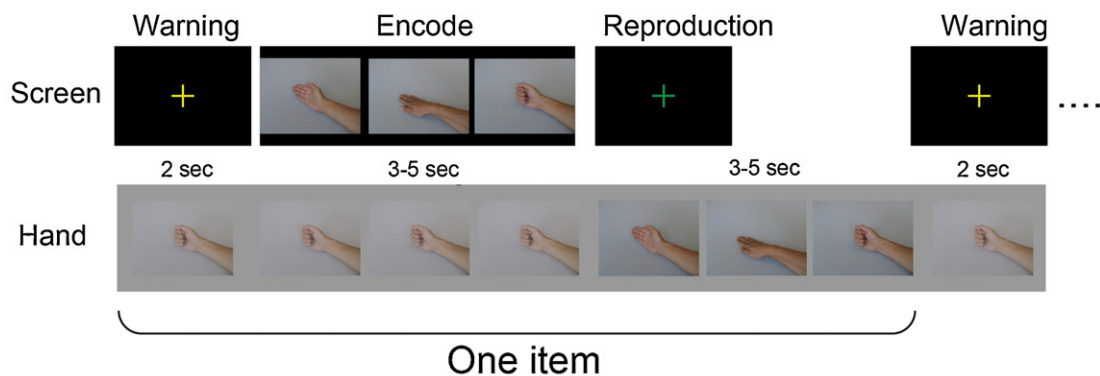Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.neures.2012.12.007.
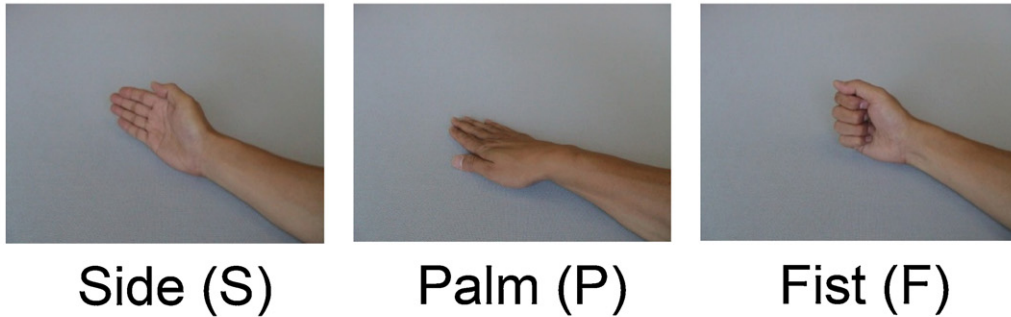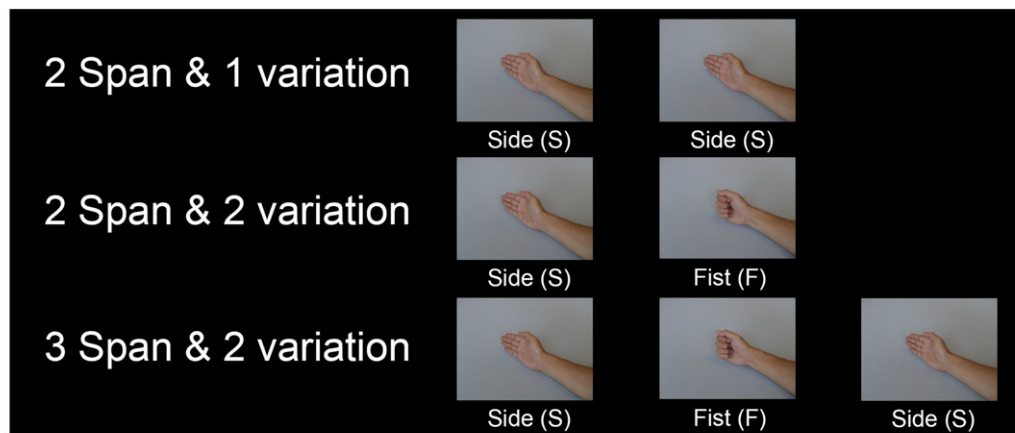
Fig. 1. Representative items from the Okazaki Hand Movement Test for the fMRI study. (A) Three hand postures were used to produce the sequences of hand movements for immediate imitation. Note that the participant made these gestures from an initial fist position. (B) Each item consisted of different spans (i.e., numbers of hand gestures) and different variations (i.e., types of hand gestures). Six different levels of complexity of hand movement were produced in terms of span and variation (see also Table 1). (C) An example of a trial. Each trial started with a yellow fixation cross. Subsequently, an item was presented on the screen. As soon as the visual presentation of the item was complete, a green fixation cross appeared. The participant reproduced the observed hand gestures over a period with the same duration as the visual observation.

All visual stimuli were generated by a personal computer (Dimension 9250; Dell Computer Co., TX), which also controlled the timing of stimulus presentation using Presentation 0.90 software (Neurobehavioral Systems, CA). The visual stimuli were back-projected from a liquid-crystal display (LCD) projector (DLA-M200L; Victor, Yokohama, Japan) onto a translucent screen, which each participant viewed through an angled mirror attached to the head coil.

2.4. Task design

We conducted one fMRI run for each participant. The 60 hand-movement items were divided into 18 task blocks, a total time of 588 s (Table 1). The number of items in each task block and the number of task blocks in each complexity level depended on the increasing sequence lengths (two to four spans) and the limited variations of hand positions (one to three variations). The

Table 1
Order of task blocks.

Task block	Complexity of hand movements		Sample	No. of items	Duration	Complexity
	Span	Variation				
1	2	1	S-S	4	32	1
2	2	1	P-P	4	32	1
3	2	1	F-F	4	32	1
4	2	2	P-F	4	32	2
5	2	2	F-S	4	32	2
6	2	2	P-S	4	32	2
7	3	2	S-F-S	3	30	3
8	3	2	P-S-P	3	30	3
9	3	3	S-P-F	3	30	4
10	3	3	F-P-S	3	30	4
11	3	3	P-F-S	3	30	4
12	3	3	F-P-S	3	30	4
13	4	2	S-F-F-S	3	36	5
14	4	2	F-P-F-P	3	36	5
15	4	3	S-P-P-F	3	36	6
16	4	3	P-F-P-S	3	36	6
17	4	3	F-P-S-F	3	36	6
18	4	3	S-P-F-P	3	36	6

Note. Each task block consists of the same type of items and 18 task blocks were presented per run, in the order of six levels of increasing complexity. Span: number of serial hand positions; Variation: the number of different types of hand positions used for the item; Sample: Example of the item; S: Side; F: Fist; P: Palm (Fig. 1A). Six different regressors were used to model these 18 blocks; assigned number in each task block indicates which regressor was used to model the task block. The unit of duration is in second.

18 task blocks therefore differed both in the number of items (three or four) and in duration (30–36 s), and the number of task blocks in the six complexity levels varied from two to four (Table 1). The numbers of items and task blocks, and the durations were inconsistent because we were particularly interested to examine the effect of spontaneous formation of verbal labels with increasing complexity levels. We inserted a 16-s baseline period before the first task block and a 12-s baseline period after each task block (16 s + 12 s × 18 task blocks = 232 s). Altogether, the total run time was 820 s (232 s + 588 s, 410 images).

Instructions, practice trials and a quiz to confirm the understanding of the instructions were administered outside the MRI scan room. In accordance with the Kaufman Hand Movement Test (KHMT) (Kaufman and Kaufman, 1983), the practice trials consisted of the imitation of the three position one after another and a Palm-Palm item. The participants were only asked to accurately imitate hand movements at the same speed, and no information was given about verbal labels. During the baseline period of the fMRI experiment, a white fixation cross was presented against a black background at the center of the display frame. An onset of each trial within a task block started with the appearance of a yellow fixation cross for 2 s as a warning sign (Fig. 1C). The yellow cross was then replaced by a video clip demonstrating a hand-movement sequence from 3 to 5 s. In all sequential video clips, the hand in the fist position (home position) started moving at the same time; the hand was displaced in the second frame at standard analog NTSC video rates (29.97 frames per second) toward the first position where the hand arrived at the first second from the start of the video clip. The duration of each video clip depended on the span of the movement, ranging from two to four. Each participant was instructed to keep his or her right hand still, in the fist position (home position), on a transparent table that was placed over the stomach area, while awaiting and watching the hand movement. After viewing the hand movement, a green fixation cross appeared on the display as a signal for the participant to immediately reproduce the demonstrated hand movements in the same sequential order and at the same speed. The green fixation cross was presented for the same time period as the video clip. It was then replaced by a yellow warning sign which indicated to the participant that the video clip for the next item was about to be shown. The duration

of a trial starts with the yellow fixation cross and ends with the disappearance of the green fixation cross.

2.5. Measurement of behavioral performance

A digital video camera (Handycam, HDR-SR1; Sony, Tokyo, Japan) was used to record each participant's hand movements on the transparent table through a waveguide in the wall of the MRI scanner room. The images from two video sources, one for the visual stimuli and the other for the participant's hand movements, were combined with a splitter, and simultaneously recorded side-by-side onto digital video cassette tapes (GV-D1000; Sony, Tokyo, Japan). These were later used to analyze reaction time, response time, and performance error in sequential order. Reaction time was defined as the time interval between the appearance of the green cross and the onset of hand-movement in each trial, whereas response time was a sum of reaction time plus movement time required until the hand arrived at the first, second, third or the fourth position. If a participant started to move his or her hand while watching a video clip before the start signal, the trial was excluded from the data analysis of reaction time, response time, and sequential error.

2.6. fMRI data processing

Image processing and statistical analyses were performed with the Statistical Parametric Mapping package (SPM8; The Wellcome Trust Centre for Neuroimaging, London, UK) implemented in Matlab (Mathworks, Natick, MA). The first five volumes (i.e., the initial 10 s waiting time) of each fMRI session were discarded to allow the MR signal to reach a state of equilibrium, and the remaining 405 volumes per session were used for analysis. All functional images and the T1-weighted anatomical images were then co-registered to the first scan of the hand-movement encoding and imitation task. Each co-registered T1-weighted anatomical image was normalized to a standard T1 template image (ICBM 152), which defined the Montreal Neurological Institute (MNI) space. The parameters from this normalization process were then applied to each functional image. We spatially filtered the normalized echo-planar images, using a

Table 2
Predefined contrasts for the random-effect group analysis.

Block (complexity)	Regressors											
	1		2		3		4		5		6	
Verbal strategy	VL−	VL+	VL−	VL+	VL−	VL+	VL−	VL+	VL−	VL+	VL−	VL+
VL− vs. baseline	1	0	1	0	1	0	1	0	1	0	1	0
VL+ vs. baseline	0	1	0	1	0	1	0	1	0	1	0	1
VL− vs. VL+	1	−1	1	−1	1	−1	1	−1	1	−1	1	−1
VL+ vs. VL−	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1

Note. VL−: strategy without the use of verbal labels; VL+: strategy with the use of verbal labels.

Gaussian kernel of 8 mm full-width at half maximum (FWHM) in the *x*, *y*, and *z* axes.

2.6.1. First-level fMRI analysis

We fitted a general linear model to the functional MRI data for each participant (Friston et al., 1994; Worsley and Friston, 1995), and prepared one design matrix for each participant. Each task block consisted of items of the same complexity, and all the blocks in a run could be classified into six different levels of complexity (Table 1). We therefore introduced six different regressors with each reflecting respective levels of complexity in order to model the 18 task blocks in the run. The neural activity during each task block was modeled with box-car functions convolved with the canonical hemodynamic-response function. The time series for each voxel was high-pass filtered at 1/128 Hz. Assuming a first-order autoregressive model, the serial autocorrelation was estimated from the pooled active voxels with the restricted maximum likelihood (ReML) procedure, and was used to whiten the data and design matrix (Friston et al., 2002). Motion-related artifacts were minimized by incorporating six parameters (three displacements and three rotations) from the rigid-body realignment stage into each model. We evaluated the estimates for conditions with six levels of complexity, using linear contrasts. The condition of each level of complexity, consisting of both observation and reproducing hand movements, was compared with the baseline. The observation and immediate imitation periods were modeled together because the involvement of both encoding and retrieval processes during the observation period made it impossible to distinctively assign the observation period to encoding and the imitation period to retrieval. The comparison produced a contrast image for each participant, and these contrast images were used for the second-level fMRI analysis.

2.6.2. Second-level fMRI analysis

Contrast images from the individual analyses were used for the group analysis, with between-subjects variance modeled as a random factor (random-effects analysis). The contrast images obtained from the individual analyses represented the normalized task-related increment of the MR signal for each participant.

We constructed a single design matrix to examine brain activation corresponding to different VL strategies. Contrast images obtained at the single-subject level were classified either as using the VL strategy (VL+) or not (VL-) for each of the six complexity levels. This categorization was determined based on individual postdiction (Light, 1991) outside the MRI scanner (i.e., retrospective self-reporting) of the exact complexity level at which each participant started to use verbal labels in a questionnaire consisting of a forced-choice question about the starting point and open questions about the labels used for each hand position: all the items in the complexity levels below that reported as the starting point were considered to be VL- trials, and all of the items in the reported complexity level and those above it were considered to be VL+ trials.

A second-level design matrix was produced by defining the following two factors: first, VL user vs. non-user; and second, six levels of increasing movement complexity (12 regressors in total; Table 2). Complexity was a within-subject factor, whereas VL could be either a within-subject factor or a between-subject factor, as the participants could change their VL strategy. We considered the VL strategy to be a between-subject factor, assuming that the mixed factorial design was more conservative than a model in which the VL strategy was included as a within-subject factor. The experimental design might therefore have been unbalanced in terms of VL+/VL- and complexity level, depending on the numbers of participants using VL at each complexity level. However, we predicted that treating VL as an independent factor in our analysis would compensate for Type I error. We compared the estimates for the conditions with different levels of movement complexities, using linear contrasts. To examine the group effect of VL (i.e., verbal label users and non-users), random effect was used as an analysis of choice because it could afford population-level inferences. The resulting set of voxel values for each contrast constituted the SPM, the threshold for which was set at $p < .05$ family-wise error (FWE) rate corrected for multiple comparisons over the search volume. Brain regions were anatomically defined and labeled according to the probabilistic brain atlas (Shattuck et al., 2008).

Table 2 shows the linear contrasts employed in the present study. We initially evaluated the brain activation for each strategy relative to baseline ("VL- vs. baseline" and "VL+ vs. baseline", respectively). The search volume was the whole brain. Subsequently, we evaluated the difference between VL- and VL+ (VL- vs. VL+ and VL+ vs. VL-). The search volume for VL- vs. VL+ was limited to the regions depicted by the contrast of VL- vs. baseline $(83,304\,\text{mm}^3)$. Similarly, the search volume for VL+ vs. VL− was limited to the regions depicted by the contrast of VL+ vs. baseline $(125,352\,\text{mm}^3)$.

3. Results

3.1. Behavioral data

3.1.1. Error rate and timing of VL formation

The mean ± standard deviation (SD) number of sequential errors made by the 18 participants was 2.50 ± 3.00, and the mean error rate was 4%. Two of the participants made more errors than the mean + 2SD, and were therefore excluded from the analysis. Data from the remaining 16 participants were analyzed. The numbers of participants who started to use VL, such as "fist", "side", "flat" in each complexity level are shown in Fig. 2A. It is theoretically plausible to label a part of each hand sequence. But in practice, none of the participants in this study reported partial labeling. The number of VL users increased gradually with movement complexity as the run progressed. Overall, the strategy involving VL was employed in 56% of the trials among all of the participants (VL+ condition), whereas the alternative strategy without VL was

employed in the rest of the trials (Table 3). We investigated the strength of association between VL and complexity by computing $\eta = 0.33$. This moderate association suggested the violation of the independence assumption for factorial design analysis. We still conducted factorial ANOVA for behavioral data, and interpreted the results with caution. The violation of the assumption led to more serious threat to validity of the results obtained from the factorial design analysis of fMRI data. Therefore, we limited our primary report to subtraction analysis between VL+ and VL− conditions that examined the effect of VL without considering complexity.
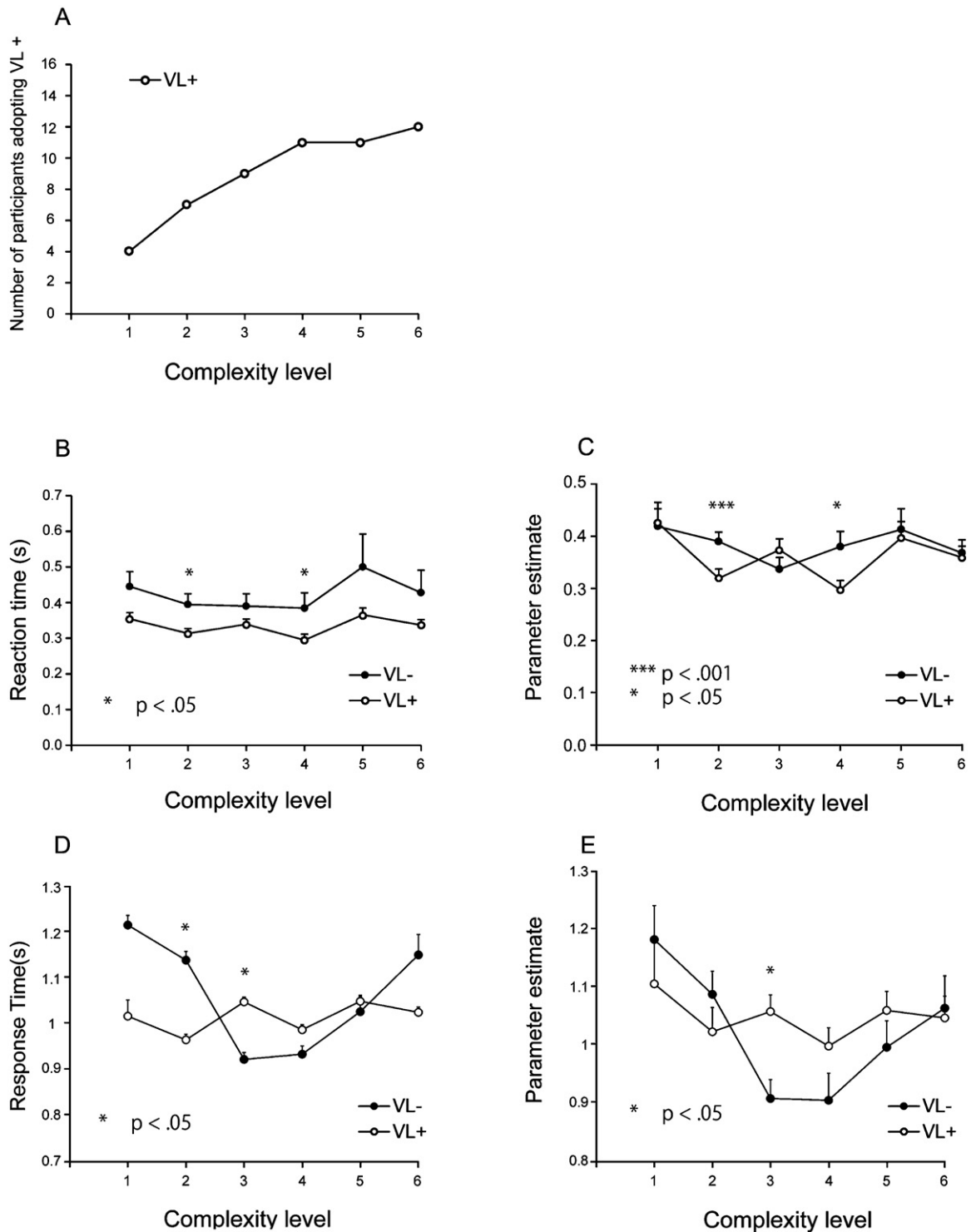
Fig. 2. Behavioral performance. (A) Cumulative frequency of verbal label users (VL+) over six levels. (B) Raw data for reaction time. (C) Parameter estimates for reaction time. (D) Raw data for response time to the first position. (E) Parameter estimates for response time to the first position. (F) Raw data for response time to the second position. (G) Parameter estimates for response time to the second position. (H) Raw data for response time to the third position. (I) Parameter estimates for response time to the third position. (J) Raw data for response time to the fourth position. (K) Parameter estimates for response time to the fourth position. (L) Raw data for error rate. (M) Parameter estimates for error rate. Significant time differences between (VL−) and (VL+) conditions at respective complexity levels are indicated (*$p < .05$, **$p < .01$). Error bar indicates standard error of the mean (SEM).

3.1.2. Reaction time, response time, and error rate between VL+ and VL− trials

Reaction time and response time data from ~3% (33 out of 960) of the trials were excluded from the analysis because the participants started to move their hands while watching the video clips. For one participant, reaction time data were unavailable due to a technical problem with the video-recording system. We did not include both reaction time and error rate data from this participant in the behavioral data analysis; however, the fMRI data were available, and were included in the analysis.

Fig. 2B–M shows the raw data and the parametric estimates for reaction time, response time, and error rate under the VL+ and VL− conditions. We presented the parameter estimates because we employed a mixed linear model in the analysis of behavioral data to emulate the procedure for the fMRI data analysis. We performed two-way analysis of variance (ANOVA) with VL (VL+ vs. VL−) serving as an independent factor and complexity (six levels) as dependent factor on reaction time, response time, and error rate.

There was a significant interaction between VL and complexity ($F_{(5,16.1)} = 5.1$, $p < .01$) on reaction time. The main effect was
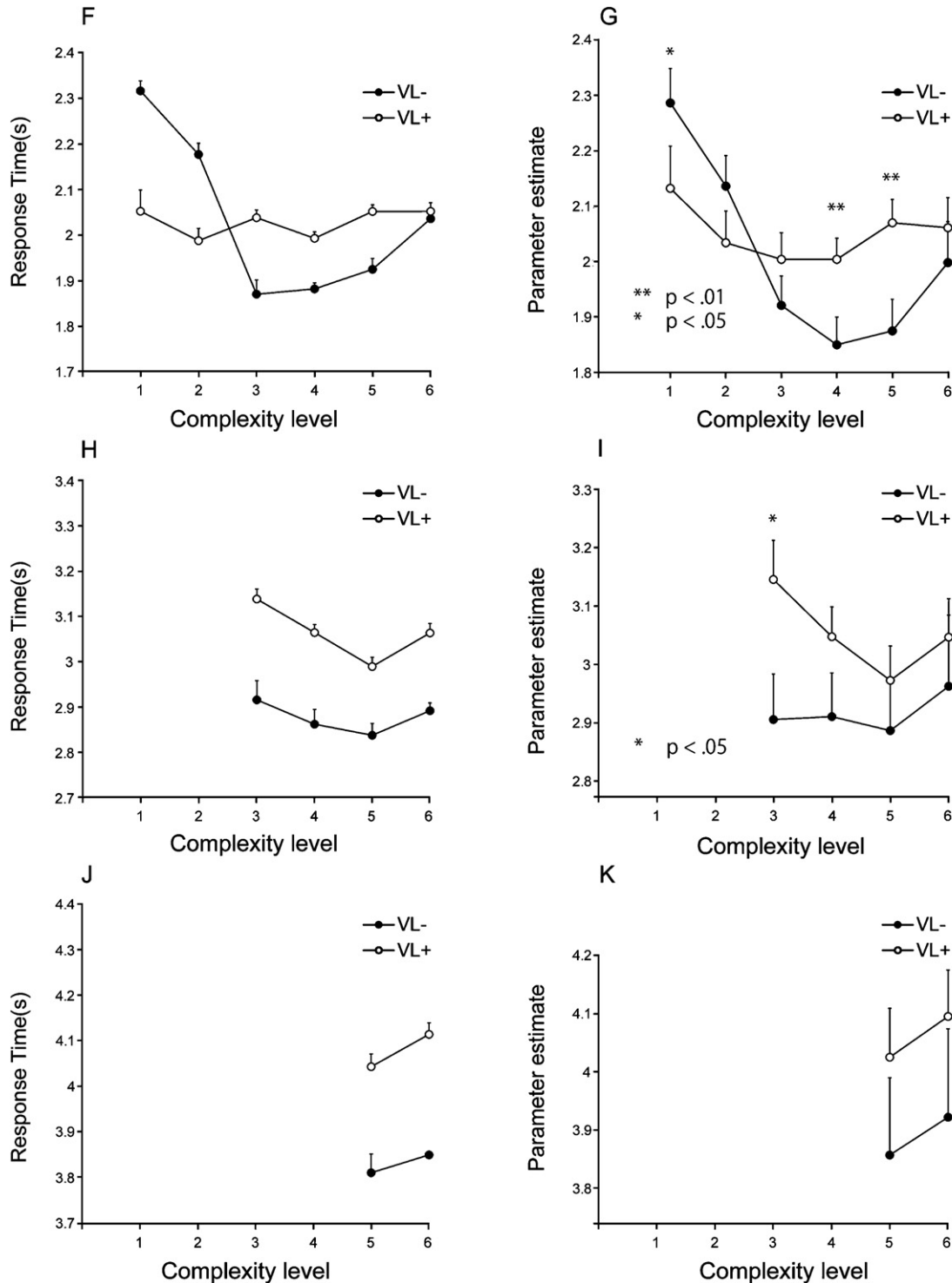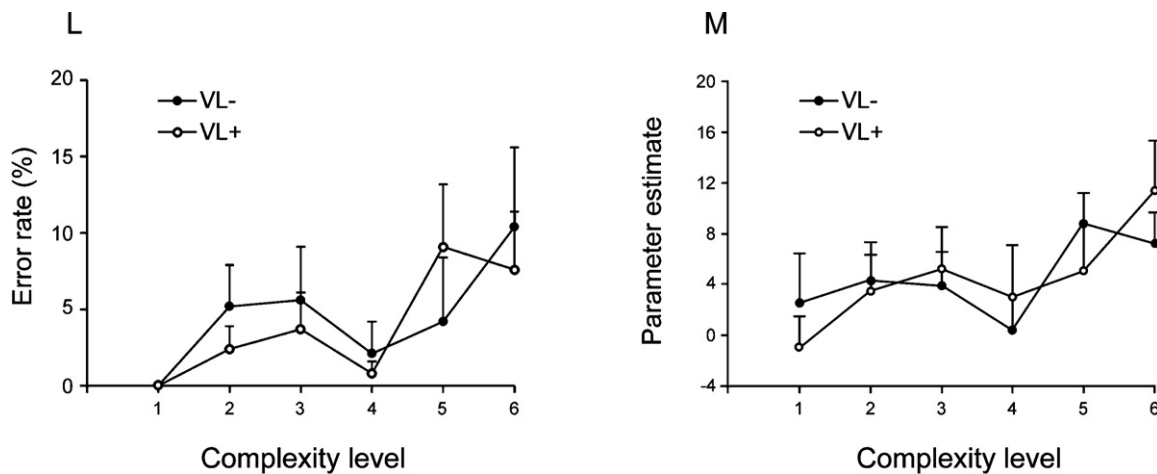

Fig. 2. (*Continued*)

Fig. 2. (*Continued*).

Table 3
The numbers of VL+ and VL− in each complexity level.

Complexity level	VL+	VL−
1	4	12
2	7	9
3	9	7
4	11	5
5	11	5
6	12	4
Subtotal	54	42

significant for complexity ($F_{(5,13.1)} = 3.3$, $p < .05$), but not significant for VL ($p > .09$). Multiple comparisons between VL+ and VL− in each of the six complexity levels revealed significantly faster reaction time for VL+ than VL− at the second ($p < .001$) and forth ($p < .05$) complexity levels (Fig. 2C).

The interactions between VL and complexity were also significant for response time to the first position ($F_{(5,15.2)} = 4.7$, $p < .01$) and the second position ($F_{(5,14.6)} = 5.4$, $p < .01$). The main effect of complexity was significant for response time to the first position ($F_{(5,14.3)} = 3.4$, $p < .05$) and the second position ($F_{(5,14.1)} = 8.9$, $p < .01$), but the main effect of VL was not significant ($p > .09$). Multiple comparisons were performed between VL+ and VL− in each of the six complexity levels for response time to the first, second, third, and fourth positions. Response time to the first position was significantly slower for VL+ than VL− at the third complexity levels ($p < .05$) (Fig. 2E). While response time to the second position was

significantly faster for VL+ than VL− at the first ($p < .05$) complexity level, it was significantly slower for VL+ than VL− at the fourth and the fifth complexity levels ($p < .01$) (Fig. 2G). Response time to the third position was also significantly slower for VL+ than VL− at the third complexity level ($p < .05$) (Fig. 2I).

Error rate was less than 11% in all conditions (i.e., above 89% performance accuracy). The same two-way ANOVA on error rate yielded neither significant main effect nor interaction (p values > .06) (Fig. 2M).

3.2. fMRI data analysis

3.2.1. Brain regions activated by imitation of hand movements relative to baseline

Fig. 3 shows the brain regions that were activated by observing and reproducing hand movements with and without verbal labels (VL− vs. baseline and VL+ vs. baseline). Whether VL was used or not, we found significant activation in bilateral brain regions encompassing the parietal (MNI coordinates: $x = -48$, $y = -38$, $z = 54$, $T_{(1,84)} = 9.70$ from VL− vs. baseline and $T_{(1,84)} = 13.30$ from VL+ vs. baseline), temporal (MNI coordinates: $x = 54$, $y = -50$, $z = 4$, $T_{(1,84)} = 5.64$ from VL− vs. baseline; $T_{(1,84)} = 5.63$ for VL+ vs. baseline), and occipital cortices (MNI coordinates: $x = -10$, $y = -94$, $z = 10$, $T_{(1,84)} = 5.27$ from VL− vs. baseline; $T_{(1,84)} = 5.31$ from VL+ vs. baseline), as well as subcortical structures such as the cerebellum (MNI coordinates: $x = 24$, $y = -52$, $z = -22$, $T_{(1,84)} = 6.62$ from VL− vs. baseline; $T_{(1,84)} = 10.05$ from VL+ vs. baseline). Conjunction
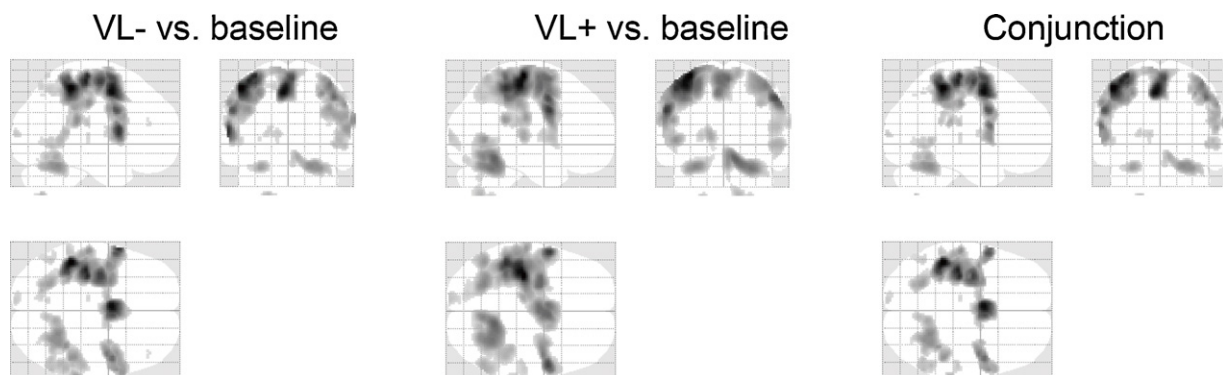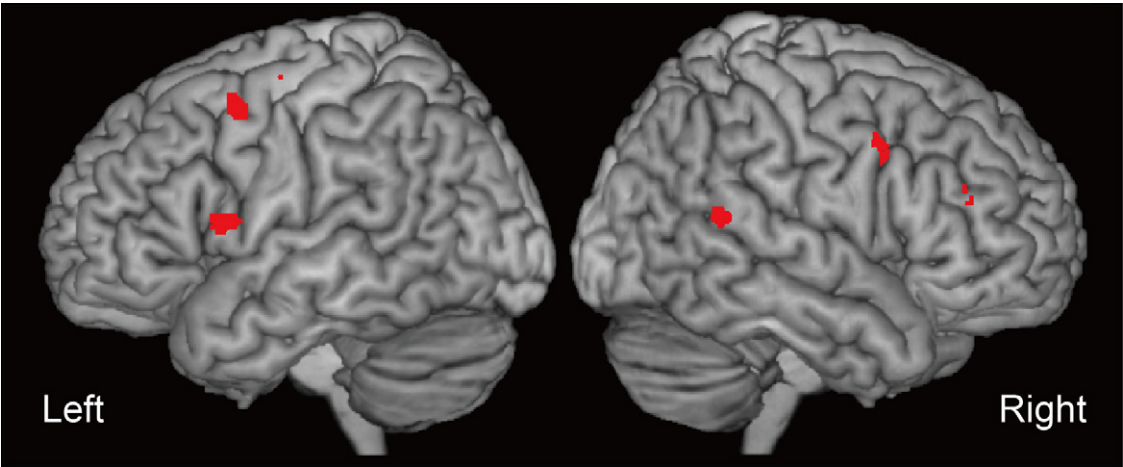

Fig. 3. Brain activation for hand imitation with each verbal strategy. VL− vs. baseline and VL+ vs. baseline shows the brain regions activated, regardless of verbal strategy (conjunction analysis). The three-dimensional (3D) information was collapsed into two-dimensional (2D) sagittal, coronal, and transverse images (i.e., maximum-intensity projections viewed from the right, back, and top of the brain). The height threshold was set at $T_{1,84} > 4.80$, equivalent to $p < .05$ FWE corrected for multiple comparisons over the whole brain.

Table 4
Brain regions activated by hand imitation without the use of the VL strategy compared to with its use (VL− vs. VL+).

Cluster size (mm³)	MNI coordinate			Hem	T value	Anatomical region
	x	y	z			
VL− vs. VL+						
184	−60	10	12	L	4.60	Inferior frontal gyrus/precentral gyrus (BA44 50%, BA45 10%)
656	−36	8	52	L	4.92	Middle frontal gyrus
960	38	6	34	R	4.85	Middle frontal gyrus (BA44 20%)
24	42	42	20	R	4.30	Middle frontal gyrus
328	−40	0	34	L	4.57	Precentral gyrus
8	−34	−10	62	L	4.07	Precentral gyrus (BA6 60%)
232	56	−44	14	R	4.50	Angular gyrus
24	40	40	22	R	4.02	Middle frontal gyrus
VL+ vs. VL−						
n.s.						

Note. The height threshold was set at $T = 4.00$ (equivalent to $p < 0.05$ FWE corrected for multiple comparisons over the brain regions activated by VL− vs. baseline). Hem: hemisphere; R: right; L: left; BA: Brodmann area based on the cytoarchitectonic brain atlas developed by Zilles and colleagues (Amunts et al., 1999); n.s.: no significantly activated region was observed.

Fig. 4. Brain regions activated by hand imitation without verbal strategy as compared to hand imitation with verbal strategy (VL− vs. VL+). The activation patterns depicted by VL− vs. VL+ are shown on the surface-rendered T1-weighted high-resolution MRI of a single participant unrelated to this study. The height threshold was set at $T_{1,84} > 4.00$, equivalent to $p < .05$ FWE corrected for multiple comparisons over the search volume (i.e., the brain regions activated by VL− vs. baseline). No significant activation was observed in the opposite contrast (VL+ vs. VL−).

analysis with the null hypothesis that one or more of the effects is null (Friston et al., 2005) confirmed significant activation in the middle temporal gyrus (MNI coordinates: $x = 54$, $y = −50$, $z = 4$, $T_{(1,84)} = 5.63$), inferior parietal lobule (MNI coordinates: $x = −48$, $y = −38$, $z = 54$, $T_{(1,84)} = 9.70$), and IFG in the bilateral hemispheres. Probabilistic cytoarchitectonic maps of Brodmann area (BA) 44 and BA45 (Amunts et al., 1999) revealed that a peak coordinate in the left IFG (MNI coordinates: $x = −56$, $y = 10$, $z = 6$, $T_{(1,84)} = 6.41$) might be located within BA44 (with 40% probability) or BA45 (with 10% probability), whereas the peak coordinate in the right IFG ($x = 58$, $y = 14$, $z = 28$, $T_{(1,84)} = 5.98$) might be located within BA44 (with 70% probability). This result confirmed that the OHMT activated the bilateral IFG adjacent to BA44 and BA45, whether VL was used or not.

3.2.2. Brain activation produced by different VL strategies

Table 4 and Fig. 4 show brain regions that produced significantly greater activation when VL was not used (VL− condition) compared to when it was used (VL+ condition). We performed this analysis, using an inclusive masking procedure in which the statistical maps for (VL− vs. VL+) condition were inclusively masked by those for the (VL− vs. baseline) condition. Thus analyzed contrast of VL− vs. VL+ revealed regions of significant activation in the left IFG, bilateral middle frontal gyrus (MFG), left precentral gyrus (PCG), and right angular gyrus (AG). A peak coordinate in the left IFG might have

been located in BA44 (with 50% probability) (Amunts et al., 1999), and a peak coordinate in the left PCG might have been located in BA6 (with 60% probability) (Geyer, 2003). Conversely, we observed no significant activation when the VL+ condition was compared with the VL− condition.

4. Discussion

This study identified the neural substrates involved in the effect of using VL on sequential working memory for hand movements. We hypothesized that VL would lower neuronal activity in the imitation network for the working memory. The behavioral data provided partial evidence for an effect of VL on reaction and response times. The fMRI data supported the hypothesis on activation in the left IFG as well as other imitation-related areas, i.e., the bilateral middle frontal gyri, the left precentral gyrus, and the right angular gyrus.

Alongside a function in motor speech, previous studies have revealed involvement of the IFG in decoding, understanding, and associating gestures with verbal language (Ackermann et al., 2004; Nishitani et al., 2005; Fadiga and Craighero, 2006; Willems et al., 2006), in covert articulation (Paulesu et al., 1993) and in sequence processing (Schubotz and Fiebach, 2006; Bahlmann et al., 2009). Our present results suggest that visual processing of hand movements in the IFG is more costly than processing the same gesture

sequences using verbal labels. The extra cost of the brain activation may reflect the price that verbal label non-users have to pay for the absence of perceived self efficacy, as well as for the taxing memory process. Whatever the components of the extra cost may be, any neural activity cost in the IFG associated with using the VL strategy is smaller than the benefit gained by employing verbal labels (i.e., the saving compared with the cost that would have been incurred without VL). This interpretation assumes that two modalities of sequence observation and imitation share a network centered in the IFG.

In addition to the left IFG, the second-level fMRI analyses commonly identified activation in the middle frontal gyrus (MFG), the left precentral gyrus (PCG), and the right angular gyrus (AG) as contributing to the cost of not using verbal labels. Previous studies showed that these regions were involved in the processing of gesture recognition (Bavelier et al., 2008; Fadiga et al., 2006), sequence learning (Kawashima et al., 1995; Müller et al., 2002), sequence processing (Bahlmann et al., 2009), and visual working memory (Gruber and Von Cramon, 2003; Ventre-Dominey et al., 2005; Moore et al., 2006). These are crucial functional components of the immediate imitation task without VL. Consistent with our hypothesis, greater activation was observed in this imitation network.

If verbal labels are not used, each element of the sequential hand movements must be visually encoded and retrieved, as is the case when a novice learns a new skill without automatizing the sequential procedure of its components. Previous research reported that the right MFG (Müller et al., 2002) and the right AG (Kawashima et al., 1995) were activated only in the early stage of motor learning. The early stage of learning does not involve automatization, whereas the late stage does; activation in these two regions at the early stage of learning might therefore be related to the absence of automatization. As we argued in the introduction, VL can be considered as partial automatization. The right MFG and the right AG of those participants who did not use verbal labels or partial automatization might have been continuously activated in a way similar to a novice during the early stage of learning.

We analyzed fMRI data during task blocks in which observation and immediate imitation of sequential hand movements occurred. In this method, it is unclear which phase of the memory process — encoding, retention, or retrieval — carries the extra activation cost when VL is not used because we analyzed the activation in both observation and imitation phases together. This may be also why activity in the occipital and parietal cortices, which have been deemed to be involved in visual imagery (Munzert et al., 2009) and visuo-spatial processing (Kravitz et al., 2011), was not significant when the VL+ condition was compared with the VL− condition in the second-level fMRI analyses, although the first-level fMRI analysis identified activation in these areas. Perhaps, the observation of hand movements activated the same areas equally, and the activation of the areas during the imitation without VL might be negligible compared to the activity during the observation phase. To examine phase-specific activation in the IFG and other regions involved in visual imagery, visual processing without VLs and covert articulation, future research will need to employ a certain task design that separates the three phases of the memory process distinctively during delayed imitation tasks, and confirms the relative activation required for each strategy in different phases.

In the present study we determined the use of VL on the basis of postdiction, or retrospective self-reporting, the validity of which was confirmed by the consistent motor span (approximately three) of sequential hand movements whether the motor span was inferred by articulatory suppression method (Frencham et al., 2004) or the postdiction method (Miyahara, 2007), and also in part by our reaction time, response time, and fMRI data. Reaction time tended to be faster in the VL+ conditions than the

VL− conditions, presumably because verbal labeling allowed faster retrieval of motor chunks compared with VL non-users who must recollect a visual memory of the same sequences, analogous to re-playing a video clip. Response time, or the sum of reaction time plus movement time, by contrast, was often slower in the VL+ conditions than the VL− conditions with increasingly complex sequences. A possible explanation for this could be that VL users perceived themselves to be more efficacious, confident and relaxed in reproducing the sequences than the VL non-users who would feel pressured to finish off the sequences before they faded or were lost from their visual working memory.

The possible reasons for the incomplete support for the efficiency of VL from the reaction time and response time data include the wide individual variation in the small number of participants, our instruction which did not specifically ask the participants to reproduce the hand movements as soon as possible, but immediately after the cue, and a possibility of switching back and forth between the visual and VL strategies. Indeed, some participants may have employed strategies other than the visual or the verbal strategy to perform the task. Such strategies, which might not be specifically related to language processing, could have not been captured in the forced-choice structure of the retrospective report. This may explain the lack of language-specific activation when the VL+ condition was compared with the VL− condition.

Our study is also limited by the violation of the independence assumption for the factorial design because of our priority for spontaneous formation of verbal labels at the cost of losing experimental control. Therefore, it may be argued that VL dependent deactivation in the IFG and other regions could in fact reflect complexity. We have counter arguments supported by some evidence. Previous behavioral studies (Frencham et al., 2003; Miyahara, 2007) provided data that VL expanded the capacity of visuospatial memory. Past fMRI studies indicated the involvement of Broca's area in processing syntactic sequences (Fiebach and Schubotz, 2006), and suggested that the IFG was more strongly activated by processing complex sequences than simple ones (Fiebach et al., 2005; Bahlmann et al., 2009). Another line of fMRI research demonstrated that less brain activity was used to perform sequential movements in the late automatic stage of sequence learning than in the initial stage (Sadato et al., 1996; Wu et al., 2004). Although our task is essentially one-trial learning, not a well learned task by repetition (cf. Sadato et al., 1996; Wu et al., 2004), verbal labeling of hand movements still makes the transition process from one position to another "automatic" by shortcutting the visual encoding and retrieval of entire continuous hand movements. Taken together, it is more likely that increased complexity should lead to greater activation in the imitation-related regions, not deactivation.

Moreover, we have supplementary data which evaluated brain regions where increment of activity related to complexity levels was greater for VL− condition than VL+ condition. In other words, the slope of the VL− regression line is steeper than the slope of the VL+ regression line (see Supplementary Data, Fig. S1). This analysis is also vulnerable to the same criticism on the violation of independence assumption, hence a factorial design analysis is not suitable. Nevertheless, this analysis indicates that VL dependent activation (VL+) in the IFG does not reflect complexity, whereas VL− does (Table S1) — a possibility that VL may make the activation in the imitation-related regions remain constant regardless of movement complexity.

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.neures.2012.12.007.

In concert with foregoing studies, our results suggested that VL of hand movements might suppress activation in the IFG and other imitation network. VL seems to make an effect of increasing neural efficiency by expanding visuospatial memory capacity, while

reducing cortical activity. This preliminary finding needs to be verified by future research with stringent experiment control over VL. We plan to perform a subsequent study by preventing participants from using verbal labels with the articulatory suppression method (Baddeley, 1986; Frencham et al., 2004).

References

Ackermann, H., Mathiak, K., Ivry, R.B., 2004. Temporal organization of "internal speech" as a basis for cerebellar modulation of cognitive functions. Behav. Cognit. Neurosci. Rev. 3, 14–22.
Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H.B., Zilles, K., 1999. Broca's region revisited: cytoarchitecture and intersubject variability. J. Comp. Neurol. 412, 319–341.
Baddeley, A., 1986. Working Memory. Clarendon Press, Oxford.
Baddeley, A., Lewis, V., Vallar, G., 1984. Exploring the articulatory loop. Q. J. Exp. Psychol. A 36, 233–252.
Bahlmann, J.G., Schubotz, R.I., Mueller, J.L., Koester, D., Friederici, A.D., 2009. Neural circuits of hierarchical visuo-spatial sequence processing. Brain Res. 1298, 161–170.
Barkley, R.A., 2004. Attention-deficit/hyperactivity disorder and self-regulation: taking an evolutionary perspective on executive functioning. In: Baumeister, R.F., Vohs, K.D. (Eds.), Handbook of Self-Regulation: Research, Theory, and Applications. Guilford Press, New York.
Bavelier, D., Newman, A.J., Mukherjee, M., Hauser, P., Kemeny, S., Braun, A., Boutla, M., 2008. Endocing, rehearssal, and recall in signers and speakers: shared network but differential engagement. Cereb. Cortex 18, 2263–2274.
Berkowitz, A.L., Ansari, D., 2008. Generation of novel motor sequences: the neural correlates of musical improvisation. NeuroImage 41, 535–543.
Caspers, S., Zilles, K., Laird, A.R., Eickhoff, S.B., 2010. ALE meta-analysis of action observation and imitation in the human brain. Neuroimage 50, 1148–1167.
Fadiga, L., Craighero, L., 2006. Hand actions and speech repreesntation in Broca's area. Cortex 42, 486–490.
Fadiga, L., Craighero, L., Roy, A., 2006. Broca's region: a speech area? In: Grodzinsky, Y., Amunts, K. (Eds.), Broca's Region. Oxford Press, Oxford.
Fiebach, C.J., Schlesewsky, M., Lohmann, G., Von Cramon, D., Friederici, A., 2005. Revisiting the role of Broca's area in sentence processing: syntactic integration versus syntactic working memory. Hum. Brain Mapp. 24, 79–91.
Fiebach, C.J., Schubotz, R.I., 2006. Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains? Cortex 42, 499–502.
Fitts, P.M., 1964. Perceptual-motor skill learning. In: Melton, A.W. (Ed.), Categories of Human Learning. Academic Press, New York.
Frencham, K.A.R., Fox, A.M., Maybery, M.T., 2003. The Hand Movement Test as a tool in neuropsychological assessment: interpretation within a working memory theoretical framework. J. Int. Neuropsych. Soc. 9, 633–641.
Frencham, K.A.R., Fox, A.M., Maybery, M.T., 2004. Effects of verbal labeling on memory for hand movements. J. Int. Neuropsych. Soc. 10, 355–361.
Friston, K.J., Glaser, D.E., Henson, R.N., Kiebel, S., Phillips, C., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: applications. NeuroImage 16, 484–512.
Friston, K.J., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time-series. Hum. Brain Mapp. 1, 153–171.
Friston, K.J., Penny, W.D., Glaser, D.E., 2005. Conjunction revisited. NeuroImage 25, 661–667.
Geschwind, N., 1964. The development of the brain and the evolution of language. In: Stuart, C.I.J.M. (Ed.), Monograph Series on Language and Linguistics. Georgetown University Press, Washington, D.C.
Geyer, S., 2003. The Microstructural Border Between the Motor and the Cognitive Domain in the Human Cerebral Cortex. Springer, Wien.
Gmeindl, L., Walsh, M., Courtney, S.M., 2011. Binding serial order to representations in working memory: a spatial/verbal dissociation. Memory Cognition 39, 37–46.
Gruber, O., Von Cramon, D.Y., 2003. The functional neuroanatomy of human working memory revisited: evidence from 3-T fMRI studies using classical domain-specific interference tasks. NeuroImage 19, 797–809.
Haxby, J., Petit, L., Unerleider, L., Courtney, S., 2000. Distinguishing the functional roles of multiple regions in distributed neural systems for visual working memory. NeuroImage 11, 145–156.
Hikosaka, O., Nakahara, H., Rand, M.K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., Doya, K., 1999. Parallel neural networks for learning sequential procedures. Trends Neurosci. 22, 464–471.
Kaufman, A.S., Kaufman, N.L., 1983. Kaufman Assessment Battery for Children: Administration and Scoring Manual. American Guidance Service, Circle Pines, MN.
Kawashima, R., Roland, P.E., O'sullivan, B.T., 1995. Functional anatomy of reaching and visuomotor learning: a positron emission tomography study. Cereb. Cortex 2, 111–122.
Keele, S.W., 1973. Attention and Human Performance. Goodyear, Pacific Palisades, CA.
Kravitz, D.J., Saleem, K.S., Baker, C.I., Mishkin, M., 2011. A new neural framework for visuospatial processing. Nat. Rev. Neurosci. 12, 217–230.
Light, L.L., 1991. Memory and aging: four hypotheses in search of data. Annu. Rev. Psychol. 3, 333–376.
Luria, A., 1966. Higher Cortical Functions in Man. Basic Books, New York.
Meltzoff, A.N., 1988. Imitation, objects, tools, and the rudiments of language in human ontogeny. Hum. Evol. 3, 45–64.
Miyahara, M., 2007. Aging influences on working memory for hand movements: a test of the metamemory deficit hypothesis. Exp. Aging Res. 33, 417–428.
Miyahara, M., Leeder, T., Francis, G., Inghelbrecht, A., 2008. Does an instruction of a verbal labeling strategy for hand movements improve general motor coordination as well as the gestural performance? A test of the relationship between developmental coordination disorder and dyspraxia. Clin. Case Studies 7, 191–207.
Moore, C.D., Cohen, M.X., Ranganath, C., 2006. Neural mechanisms of expert skills in visual working memory. J. Neurosci. 26, 11187–11196.
Müller, R., Kleinhans, N., Pierceb, K., Kemmotsua, N., Courchesne, E., 2002. Functional MRI of motor sequence acquisition: effects of learning stage and performance. Cognitive Brain Res. 14, 277–293.
Munzert, J., Lorey, B., Zentgraf, K., 2009. Cognitive motor processes: the role of motor imagery in the study of motor representations. Brain Res. Rev. 60, 306–326.
Nishitani, N., Schurmann, M., Amunts, K., Hari, R., 2005. Broca's region: from action to language. Physiology 20, 60–69.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh Inventory. Neuropsychologia 9, 97–113.
Owen, A.M., 2004. Working memory: imaging the magic number four. Curr. Biol. 14, R573–R574.
Paulesu, E., Frith, C.D., Frackowiak, R.S.J., 1993. The neural correlates of the verbal component of working memory. Nature 362, 342–345.
Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. Trends Neurosci. 21, 188–194.
Sadato, N., Campbell, G., Ibanez, V., Deiber, M.P., Hallett, M., 1996. Complexity affects regional cerebral blood flow change during sequential finger movements. J. Neurosci. 16, 2693–2700.
Santia, A., Grodzinskya, Y., 2007. Working memory and syntax interact in Broca's area. NeuroImage 37, 8–17.
Schmidt, R.A., Lee, T.D., 1999. Motor Control and Learning: A Behavioral Emphasis. Human Kinetics, Champaign, IL.
Schubotz, R.I., Fiebach, C.J., 2006. Integrative models of Broca's area and the ventral premotor cortex. Cortex 42, 461–463.
Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W., 2008. Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39, 1064–1080.
Takahama, S., Miyauchi, S., Saiki, J., 2010. Neural basis for dynamic updating of object representation in visual working memory. NeuroImage 49, 3394–3403.
Tettamanti, M., Rotondi, I., Perani, D., Scotti, G., Fazio, F., Cappa, S.F., Moro, A., 2009. Syntax without language: neurobiological evidence for cross-domain syntactic computations. Cortex 45, 825–838.
Ventre-Dominey, J., Bailly, A., Lavenne, F., Lebars, D., Mollion, H., Costes, N., Dominey, P.F., 2005. Double dissociation in neural correlates of visual working memory: a PET study. Cognitive Brain Res. 25, 747–759.
Willems, R.M., Ozyurek, A., Hagoort, P., 2006. When language meets action: the neural integration of gesture and speech. Cereb. Cortex 17, 2322–2333.
Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited-again. NeuroImage 2, 173–181.
Wu, T., Kansaku, K., Hallett, M., 2004. How self-initiated memorized movements become automatic: a functional MRI study. J. Neurophysiol. 91, 1690–1698.