Role of the Left Inferior Frontal Gyrus in Transforming Format Types of Action Descriptions between Stimuli and Representations

Hiroshi Shibata[1] and Kenji Ogawa[2]

[1]Department of Rehabilitation, Faculty of Medical Science and Welfare, Tohoku Bunka Gakuen University, Japan

[2]Department of Psychology, Graduate School of Humanities and Human Sciences, Hokkaido University, Japan

Corresponding author: Hiroshi Shibata. Ph.D.
Address: Department of Rehabilitation, Faculty of Medical Science and Welfare, Tohoku Bunka Gakuen University, 6-45-1 Kunimi, Aoba-ku, Sendai, Miyagi 981-8551, Japan
E-mail: hshibata@rehab.tbgu.ac.jp

Abstract

We performed a functional magnetic resonance imaging study to elucidate the process involved in the transformation of the format types of action descriptions between stimuli and representations. We independently manipulated the format types of both stimuli (visual action [Vi] vs. verbal [Ve] stimulus) and internal representations (Vi vs. Ve representation) and set four types of experimental tasks. Each participant was required to generate a Vi or Ve representation after being presented with a Vi or Ve stimulus, according to each task. Increased activity in the left inferior frontal gyrus (IFG) (Brodmann areas 44 and 45) was found in the transformation contrast: ([Vi stimulus and Ve representation] + [Ve stimulus and Vi representation]) > ([Vi stimulus and Vi representation] + [Ve stimulus and Ve representation]). This result suggests that the left IFG is involved with the transformation process and has the function of generating an internal representation in a format different from that of externally presented stimuli.

Keywords: fMRI; left inferior frontal gyrus; transformation process; verbal representation; visual action representation

1. Introduction

People can internally generate various format types of representations (such as visual, auditory, and action representations) from different stimuli formats (Pavio, 2013; Price, 2012, 2018; Price et al., 2015). For example, when individuals read a sentence describing a certain action, such as *I grasp an apple*, they can internally generate a verbal representation (e.g., speech imagery) of that sentence and generate an action representation (e.g., visual action imagery) corresponding to it. Another example is when an individual looks at a picture of a grasping action; they can internally generate both visual and verbal representations that correspond to it.

People often need to transform the format types of action descriptions in various situations; for example, they may need to do something, such as cook or assemble something, by following written instructions. At other times, people need to instruct someone verbally how to do something, such as how to cook or assemble something. This study aimed to examine the brain activities related to the internal generation of visual and verbal action representations based on visual and verbal stimuli that describe actions. In particular, we examined the transformation process between the format types of action descriptions.

1.1. Previous Research on Visual Actions and Verbal Representations of Action Descriptions

It has been reported that both visually (e.g., picture and video) and verbally (e.g., word and sentence) described action stimuli commonly activate frontal-parietal motor areas, including the primary motor area (MC), precentral gyrus (PreG), premotor cortex (PMC), inferior frontal gyrus (IFG), and inferior parietal lobule (IPL) (Aziz-Zadeh et al., 2006; Boulenger et al., 2009;

Buccino et al., 2001; Hauk et al., 2004; Iacoboni & Dapretto, 2006; Raposo et al., 2009; Rizzolatti et al., 2001; Rizzolatti & Sinigaglia, 2007, 2010; Tettamanti et al., 2005; Willems et al., 2010). Effector-specific activation (or somatotopic organization) found in action executions has also been reported for both visually and verbally described action stimuli. For example, visual (e.g., a video of a kicking action) and verbal (e.g., a written word "kick") stimuli of foot-related actions increase activity in the dorsal area of the premotor cortex (dPMC), whereas visual (e.g., a video of a grasping action) and verbal (e.g., a written word "pick") stimuli of hand-related actions increased activity in the intermediate area between the dorsal and ventral areas of premotor cortex (mid-PMC) (Aziz-Zadeh et al., 2006; Boulenger et al., 2009; Buccino et al., 2001; Hauk et al., 2004; Raposo et al., 2009; Willems et al., 2010). These findings have led to the idea that the recognition of visual and verbal stimuli describing actions involves action representations related to action execution (Barsalou, 2008; Rizzolatti et al., 2001). It should also be noted, however, that common activated areas do not necessarily imply common neural representations. Wurm and Caramazza (2019) used a multi-voxel pattern analysis (MVPA) to show that while both visual and verbal stimuli describing actions commonly increased the activity in the frontal-parietal regions, and that the corresponding activation pattern was found in these regions within each stimulus type, the corresponding activation pattern across the two modalities was found in the lateral posterior temporal cortex.

Some studies have reported that the activities in motor-related areas for the recognition of action-related verbal stimuli are modulated by differences in tasks (Kulnke et al., 2020; 2021; Martin et al., 1995; Tomasino et al., 2007; 2014; van Dam et al., 2012; Willems et al., 2010; Yang & Shu, 2014). For example, Tomasino et al. (2014) used two types of tasks (imagery and letter detection) and two types of verbal stimuli (action and abstract concepts) and found that the

imagery task for action-related verbal stimuli increased the activity in the sensorimotor cortex, including the left IFG, right IFG, left superior medial gyrus, left supramarginal gyrus, and middle cingulate cortex, as compared to the letter detection task. Tomasino and Rumiati (2013) suggested that the type of task, rather than the type of stimulus, was involved in triggering the motor simulation. Willems et al. (2010) used imagery and lexical decision tasks to determine that lexical decision tasks elicit an effector-specific activation (greater activity with manual, rather than nonmanual, action words) in the left superior frontal sulcus, whereas motor imagery tasks elicit an effector-specific activation in the left dorsal precentral sulcus. Their ROI analysis also showed that effector-specific activation was present in the PMC during both the lexical decision and motor imagery tasks, but with no overlap.

Amit et al. (2017) examined the brain mechanisms of two types of thinking, visual (imagery) and verbal (inner speech). They used picture and sentence stimuli describing various situations, but did not aim to examine the generation processes of action representations. Further, they used actions with static movements (e.g., sitting) in addition to dynamic movements (e.g., poking) as their stimuli. Their experiment consisted of two tasks (recall and perception) and two stimuli (a picture and a sentence). They analyzed the percent signal changes in linguistic ROIs (IFG, middle frontal gyrus, anterior/posterior temporal gyrus, and angular gyrus in the left hemisphere) and visual ROIs (bilateral fusiform face area, extrastriate body area [EBA]). They showed that during the recall task, the intensity in the visual ROIs was of a similar strength as the picture and sentence conditions, whereas the signal intensity in the linguistic ROIs increased more in the sentence condition than in the image condition.

In sum, previous studies have shown distinct neural activities for visual and verbal formats of action representations, along with modality-general activation. However, how the

brain can prepare different formats of action representations according to task context is not clear yet. We hypothesize that a transformation process between different formats should be at play, which has not been sufficiently investigated in the past studies. We consider that systematic control of the effects of externally presented stimulus and internally generated representations (imagery) is necessary to specify the brain regions related to the transformation of action description format types between stimuli and representations. Willems et al. (2010) and Tomasino et al. (2007, 2014) showed that motor areas, including the MC, PMC, and IFG, were related to the process of generating explicit action imagery from verbal action stimuli. This process is assumed to be involved with the transformation from a verbal format (verbal stimuli-describing actions) to a visual action format (representations of visual action). However, these studies did not distinguish between the process related to the transformation (i.e., the interaction effect between stimulus and representation types) and the process related to the generation of visual action representation (i.e., the effect of representation type).

1.2. Cognitive Process of the Transformation between Visual Action and Verbal Formats

The transformation of action description format types between stimuli and representations is assumed to involve various cognitive processes, including the categorization of visual stimuli and language processes. Therefore, we summarized the cognitive processes we expected were involved in generating visual action and verbal representations from visual (picture) and verbal (sentence) stimuli-describing actions, as shown in Fig. 1. We created this figure using the brain models for transformations between input and output modalities employed in previous studies (Lambon Ralph et al., 2016; Patterson et al., 2007; Price, 2012, 2018). We describe four pathways in this figure: ViVi, from visual action stimuli to visual action

representations; ViVe, from visual action stimuli to verbal representations; VeVi, from verbal stimuli to visual action representations; and VeVe, from verbal stimuli to verbal representations.

Many studies, including those mentioned above, have identified the brain regions related to the cognitive processes shown in Fig. 1. In all four pathways, the visual process for visual stimuli (e.g., body parts and objects in picture stimuli or letters in sentence stimuli) is thought to be involved. Within the temporal-occipital lobe, category-selective regions for visual stimuli, such as body parts, objects, and letters, have been identified (e.g., EBA) (Dehaene, Nakamura, et al., 2010; Dehaene & Cohen, 2011; Downing et al., 2001; Peelen & Downing, 2007). ViVe and VeVi involve transformations between visual action and verbal formats of action descriptions, and it has been assumed that the anterior temporal lobe stores semantic amodal representations, acting as a hub in the transformation between different forms of modality information (Lambon Ralph et al., 2016; Patterson et al., 2007; Price et al., 2012). The lateral posterior temporal cortex is also believed to store stimulus-general action information (Wurm & Caramazza, 2019, 2022). Given that ViVi and VeVe do not require transformations between visual action and verbal formats—for example, it is possible to generate verbal representations without understanding the meaning of verbal stimuli—we included the direct pathway, without mediating the semantics, from the visual sensory process to representations in ViVi and VeVe.

Previous research has indicated that visual and verbal stimuli describing actions trigger increased activity in the action-related areas (Aziz-Zadeh et al., 2006; Boulenger et al., 2009; Buccino et al., 2001; Hauk et al., 2004; Iacoboni et al., 2006; Raposo et al., 2009; Rizzolatti et al., 2001; Rizzolatti & Sinigaglia, 2007, 2010; Tettamanti et al., 2005; Willems et al., 2010). Previous research has shown that action imagery engages the frontal-parietal regions,

including the MC, PMC, IFG, and IPL (Binkofski et al., 2000; Burianova et al., 2013; Decety et al., 1994; Ehrsson et al., 2003; Filimon et al., 2007; Hanakawa, 2016; Hétu et al., 2013). In relation to the language process of verbal representations, Amit et al. (2017) found that verbal representations (sentence recall) activated the language ROIs more than visual representations (picture recall). Even though internally generating verbal representations does not involve the execution of actual speech, this process (the mental imagery of speech) is thought to involve internal motor simulation and articulatory planning (Tian & Poeppel, 2012, 2013; Tian et al., 2016). Articulatory planning is thought to recruit the frontal motor regions, including the IFG, PMC, PreG, supplementary motor area (SMA), and insula (Silva et al., 2022; Tian et al., 2016).

To dissociate the brain mechanism involved in the transformation and representation-generation processes, we used a controlled manipulation of the effects of stimulus and representation. We assumed that both ViVi and VeVi are involved in the process of generating visual action representations, while ViVe and VeVe are related to the process of generating verbal representations. Unlike ViVi and VeVe, ViVe and VeVi need to generate an internal representation in a format different from that of an externally presented stimulus according to a task requirement. Because picture stimuli do not include verbal sentence information and sentence stimuli do not include visual action information, we hypothesized that participants must generate internal representations in ViVe and VeVi under the involvement of the internal process, rather than by simply generating a copy of an externally present stimuli. Therefore, we argue that ViVe and VeVi include additional cognitive processes, as compared to ViVi and VeVe. A top-down signal is assumed to be involved in determining what type of representations to generate according to a task requirement (Lambon Ralph et al., 2016; Tomasino & Rumiati, 2013). When a task needs to generate visual representations from verbal stimuli (or verbal

representations from visual stimuli), those transformation processes might need internally driven top-down processes because the presented stimuli do not include any required representation information. We assumed that systematic control of the effects of externally presented stimulus and internally generated representations would be effective for examining the transformation process. To the best of our knowledge, no other cognitive neuroimaging studies have systematically controlled the format types of input stimulus and internal representation in one experiment.

1.3. The Current Study

In this study, we independently manipulated the format types of the input stimulus (visual action [Vi] and verbal [Ve] stimuli, manipulation by stimulus presentation) and the internal representation (Vi and Ve representations, manipulation by instructions about which kind of representation should be generated). We used sentences describing transitive hand actions (e.g., “I grasp an apple”) for the Ve stimulus and pictures corresponding to that grasping action for the Vi stimulus. In particular, we aimed to identify the brain regions involved in the transformation processes. Of the four conditions, ViVe and VeVi were required to generate a representation of any format that differed from the presented stimulus. These transformation processes were deemed involved with the interaction effects between stimulus and representation (i.e., [ViVe + VeVi] > [ViVi + VeVe]). We used an exploratory description of the activity modulation related to the interaction effect and were specifically interested in elucidating the transformation processes.

2. Method

2.1. Participants

Twenty-two healthy volunteers participated in the experiment. The Japanese version of the FLANDERS handedness questionnaire confirmed all were right-handed (Nicholls, et al., 2013; Okubo, et al., 2014). One participant was excluded for low behavioral data performance (correct answers were below the mean minus 2*SD*). The data from the remaining 21 participants (13 females and 8 males, mean ± *SD*, age 22.0 ± 1.4) were analyzed. All participants provided written informed consent before the experiment. The experiment was approved by the ethics committee of Hokkaido University.

2.2. Materials

This experiment used sentences, not words, as the Ve stimuli and manipulated three parts of the Ve/Vi stimuli (subject/agent, predicate/action, and object) to make it difficult to retain both a Vi representation and a Ve representation at the same time during the task (see the Procedure section for the details of the task). Eight sentences written in kana (Japanese syllabograms) and eight pictures, selected based on our previous study (Shibata & Ogawa, 2018), served as the Ve and Vi stimuli (Fig. 2). The sentences were prepared by manipulating the subject (*I* or *you*), predicate (*grasp* or *touch*), and object (*apple* or *orange*). The pictures were prepared by manipulating the agent (first-person perspective or second-person perspective), action (grasping or touching), and object (apple or orange). We photographed the participants' right hands as they performed grasping and touching actions for the apple and orange and used the pictures as the stimuli for the first-person perspective to clarify that the first-person perspective was that of the participants themselves.

We presented the stimuli on a 32-inch MRI-compatible liquid-crystal display

(NordicNeuroLab, Bergen, Norway) with a 1920 × 1080 screen resolution. The stimuli were controlled with Psychopy software (http://www.psychopy.org/). The participants observed the stimuli through a mirror mounted above the head coil during fMRI scanning. The display was placed 170 cm away from the mirror. Each verbal stimulus subtended visual angles of 9.7° horizontally and 0.8° vertically, while each pictorial stimulus subtended horizontal and vertical angles of 5.4°.

2.3. Procedure

The type of stimulus modality, as well as the type of representation modality, were changed between the runs to avoid making the task too complex. This resulted in 4 run (or experimental) conditions: ViVi, ViVe, VeVi, and VeVe (Fig. 3). Each run condition consisted of experimental trials and catch trials. Catch trials were inserted to ensure that the participants made an internal representation of a given modality type in the experimental trials.

ViVi: A picture (visual action) stimulus was presented first (3 s), followed by a blank screen (3 s). The participants were asked to create an internal visual action image corresponding to that picture stimulus and to keep that image in mind during the presentation of the first stimulus and blank screen (total 6 s). In the experimental trials, a fixation cross (total 9 s) appeared at the center of the screen, and the participants were asked to stop generating the internal image. The participants were not required to press buttons in the experimental trials. In the catch trials, a picture stimulus was presented as the second stimulus (3 s), followed by a fixation cross (6 s). In half of the catch trials, the same picture stimulus was presented as the second stimulus, and in the other half, a picture stimulus with a change in any agent, action, or object was presented. The participants were asked to judge whether the second stimulus was the

same or different. They were asked either to press a button with their first (or middle) finger if the picture stimulus was the same or to press the button with their middle (or first) finger if the stimulus had changed. They were asked to respond as rapidly and accurately as possible. They needed to pay attention to the agent, action, and object of the stimulus to give a correct response. Button assignment was counterbalanced between the participants.

ViVe: A picture stimulus was presented first. The participants were asked to create an internal verbal image corresponding to that picture. The participants were instructed to create an auditory verbal image in their minds but not to vocalize it. We clarified the type of internal image to be made (we asked the participants to create an auditory verbal image, not a visual kana image). In the catch trials, a sentence stimulus was presented as a second stimulus. In half of the catch trials, the sentence stimulus corresponding to the first picture stimulus was presented as the second stimulus, and in the other half, a sentence stimulus with a change in the subject, predicate, or object was presented. The other procedures were the same as for ViVi.

VeVi: A sentence (verbal) stimulus was presented first. The participants were asked to create an internal visual action image corresponding to that sentence. In the catch trials, a picture stimulus was presented as the second stimulus. In half of the catch trials, the picture stimulus corresponding to the first sentence stimulus was presented as the second stimulus, and in the other half, a picture stimulus with a change in the agent, action, or object was presented. The other procedures were the same as for ViVi.

VeVe: A sentence stimulus was presented first. The participants were asked to create an internal verbal image corresponding to that sentence. In the catch trials, a sentence stimulus was presented as a second stimulus. In half of the catch trials, the same sentence stimulus as the first one was presented as a second stimulus, and in the other half, a sentence stimulus with a

change in the subject, predicate, or object was presented. The other procedures were the same as for ViVi.

The experiment consisted of 8 runs; each run condition was repeated twice. The representation condition factor was changed between the fourth and fifth runs, and the order was counterbalanced among the participants. Each run had 16 experimental and 12 catch trials. In each run, 8 types of stimuli were used twice. Half of the catch trials (6 trials) had consistent presentations between the first and second stimuli, and the other half (6 trials) had inconsistent presentations. The changes in the agent (or subject), action (or predicate), or object in the inconsistent presentations were 2 trials (a total of 6 trials) in each run. The experimental trials and the catch trials in each run were presented in random order. Each run began with the screen describing a task summary of the upcoming run condition and was followed by a fixation cross (9 s). The stimuli were presented in a blocked design; the 6 s task period (the first stimulus presentation and blank) alternated with the 9 s rest period (fixation) in the experimental trials, and the 9 s task period alternated with the 6 s rest period in the catch trials.

We inserted a short break between the fourth and fifth runs; the participants exited the MRI room after finishing the fourth run. The participants were given a description of the task procedure for the first 2 experimental conditions (or the last 2 experimental conditions) before the first (or fifth) run and were instructed about their task outside the MRI room. These instructions explained the corresponding combinations between the visual action stimulus and the verbal stimuli. The participants then performed the 2 practice sessions for the upcoming 2 experimental conditions before the first (or fifth) run outside the MRI room. Each practice session had 8 experimental and 6 catch trials.

2.4. fMRI Data Acquisition

A 3-T Prisma scanner (Siemens, Erlangen, Germany) with a 64-channel head coil was used to acquire T2*-weighted echo-planar images. A total of 143 scans were obtained for each run in a gradient echo-planar imaging sequence. The scanning parameters were: 3000 ms repetition time (TR); 30 ms echo time (TE); 90° flip angle (FA); 192×192 mm field of view (FOV); 94×94 matrix; 36 axial slices; and 3.0 mm slice thickness with a 0.75 mm gap. A T1-weighted anatomical image was acquired with the following parameters: 2300 ms TR; 2.32 ms TE; 8° FA; FOV; 256×256 mm matrix; 192 axial slices; and 1 mm slice thickness without a gap.

2.5. fMRI Data Analysis

Image preprocessing and statistical analyses were performed using SPM12 software (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/). The first three scans of each run were discarded to allow for T1 equilibration. All functional images were corrected for slice timing, realigned to correct for movement artifacts, spatially normalized to the Montreal Neurological Institute (MNI) template with a 3 mm^3 voxel size, and smoothed using a 6 × 6 × 6 mm Gaussian kernel. Condition effects at each voxel were estimated according to the general linear model.

The 6 s period of the experimental conditions (3 s period of the first stimulus and 3 s blank period) involved the following processes: perception of the stimulus, generation of the representation, and retention of the representation. The study's main purpose was to investigate the process of generating representations; although it was difficult to discriminate the time window of the representation generation process from that of other processes because it began at stimulus presentation, we modeled the first stimulus period and the blank period separately to examine the temporal changes in brain activity. The two periods (stimulus and blank periods) of

the four experimental conditions, excluding trials with incorrect responses (responses were required only in the catch trials), were modeled as separate boxcar functions convolved with a canonical hemodynamic response function. We also modeled the catch condition and the experimental conditions with incorrectly responded trials to account for the activity related to uninterested cognitive and movement processes. The 6 realignment parameters were included as regressors to correct for signal changes due to head movement. Session effects were included to account for nuisance covariates. A high-pass filter with a 128 s cut-off period was applied to remove low-frequency noise.

We conducted a 2 × 2 ANOVA with stimulus (Vi vs. Ve stimulus) and representation (Vi vs. Ve representation) for each period. Activation areas were considered significant if they passed the $p < 0.001$ threshold, uncorrected for multiple comparisons at the voxel level, and $p < 0.05$, familywise error (FWE) corrected for multiple comparisons at the cluster level for the whole brain. The interaction contrast ([ViVe + VeVi] > [ViVi + VeVe]) was considered to reflect the transformation process between the Vi and Ve modalities. We conducted two additional types of conjunction analyses ([ViVe > ViVi] ∩ [ViVe > VeVe] and [VeVi > ViVi] ∩ [VeVi > VeVe]) to examine the ViVe and VeVi transformation effects, respectively. We examined the mean percent signal changes for the four conditions for the areas found in the contrasts associated with the manipulation of representation. Using MarsBaR software (http://marsbar.sourceforge.net/), we calculated the means for each area, and the percent signal changes were also averaged across participants. Brain images were created using MRIcroGL software (https://www.nitrc.org/projects/mricrogl)

3. Results

3.1. Behavioral Data

The mean percentage (standard error; *SE*) of correctly no responded experimental trials was 99.6% (0.1), indicating a very high percentage. In the catch trials, we counted trials with a correct button press during the presentation of a second stimulus as the correct response. The mean percentages of correct responses (*SE*) in the catch trials for the ViVi, ViVe, ViVe, and VeVe conditions were 93.7% (1.8), 91.9% (1.6), 90.3% (2.0), and 91.7% (1.8), respectively. The two-way ANOVA showed no significant main effects or interaction. The mean reaction times (*SE*) for the correct responses in the catch trials for ViVi, ViVe, VeVi, and VeVe were 1067.2 ms (59.5), 1402.0 ms (70.5), 1172.2 ms (76.3), and 1370.3 ms (77.9), respectively. The two-way ANOVA with stimulus and representation showed a significant main effect of representation, $F(1,20) = 43.75$, $p < 0.001$, and interaction, $F(1,20) = 5.71$, $p < 0.05$. The simple main effect of representation was significant in both the Vi stimulus condition, $F(1,20) = 72.83$, $p < 0.001$, and the Ve stimulus condition, $F(1,20) = 11.75$, $p < 0.005$, indicating shorter reaction times in the Vi representation than in the Ve representation condition in the two stimulus conditions. The simple main effect of the stimulus was significant only in the Vi representation condition, $F(1,20) = 11.00$, $p < 0.005$, indicating a shorter reaction time in the Vi stimulus than in the Ve stimulus condition but only in the Vi representation condition.

3.2. fMRI Data

The results of the ANOVA for the stimulus period are shown in Table 1 and Figs. 4-6. The main effect of stimulus showed greater activation mainly in the bilateral temporal, parietal, and occipital lobes in the contrast of the Vi vs. Ve stimulus ([ViVi + ViVe] > [VeVi + VeVe]) and in the bilateral occipital lobe in the contrast of the Ve vs. Vi stimulus ([VeVi + VeVe] > [ViVi +

ViVe]) (Fig. 4). The main effect of representation revealed greater activation in the medial areas (medial frontal cortex [MFC], pre-supplementary motor area (preSMA), SMA, and middle cingulate cortex [MCC]), in the bilateral PreG (including the midPMC, vPMC, MC), postcentral gyrus (PoG), and cerebellum, and in the left IFG (pars opercularis and triangularis), Sylvian-parietal-temporal region (Spt) in the contrast of the Ve vs. Vi representation ([ViVe + VeVe] > [ViVi + VeVi]) (Fig. 5). The interaction contrast of the transformation effect ([ViVe + VeVi] > [ViVi + VeVe]) revealed greater activation in the left IFG (pars opercularis and triangularis; Brodmann area [BA] 44/45) (Fig. 6a). The conjunction analyses did not show any significant differences; neither the ViVe transformation contrast ([ViVe > ViVi] ∩ [ViVe > VeVe]) nor the VeVi transformation contrast ([VeVi > ViVi] ∩ [VeVi > VeVe]) was significant, although a small number of voxels (10) survived in the ViVe transformation contrast with a threshold of $p < 0.001$, uncorrected for multiple comparisons at the voxel level. Fig. 6b shows the activation areas found in the main effect of Ve representation ([ViVe + VeVe] > [ViVi + VeVi]) (gold regions) and the transformation effect ([ViVe + VeVi] > [ViVi + VeVe]) (red region). The red region partially overlaps the gold regions (3 voxels). No significant differences were found in the contrast of the Vi vs. Ve representation ([ViVi + VeVi] > [ViVe + VeVe]) or in the other interaction contrast ([ViVi + VeVe] > [ViVe + VeVi]).

The results of the ANOVA for the blank period are shown in Table 2 and Figs. 7 and 8. The main effect of stimulus showed greater activation in the bilateral occipital lobes in the contrast of the Vi vs. Ve stimulus ([ViVi + ViVe] > [VeVi + VeVe]) and greater activation mainly in the bilateral temporal and left parietal lobes in the contrast of the Ve vs. Vi stimulus ([VeVi + VeVe] > [ViVi + ViVe]) (Fig. 7). The main effect of representation revealed greater activation in the MCC, preSMA, left dPMC, and bilateral IPL in the contrast of the Vi vs. Ve representation

([ViVi + VeVi] > [ViVe + VeVe]) (Fig. 8). No significant differences were found in the Ve vs. Vi representation contrast, the interaction contrasts, or the two types of conjunction contrasts.

4. Discussion

4.1. Behavioral Results

Catch trials were inserted to ensure that the participants made the required type of internal representation. In the catch trials, the participants were asked to judge the presence or absence of a change between a first and second stimulus; therefore, they needed to compare the content of the Vi or Ve representation generated based on a first stimulus and the content of Vi or Ve stimulus presented as a second stimulus. We predicted that if the accuracy of Vi representation and Ve representation was not affected by the difference in stimulus modality, then no significant difference in the behavioral data would be found either between ViVi and VeVi or between ViVe and VeVe. Indeed, no significant difference was detected between ViVe and VeVe for the behavioral results. By contrast, the mean reaction time was significantly shorter in ViVi than in VeVi, although the difference was not very large (approximately 100 ms). This significant result might reflect the generation and retention of a more accurate Vi representation in ViVi than in VeVi; therefore, the reaction time for the change detection in the catch trials was faster in ViVi.

A significant reaction time difference also occurred between the Vi and Ve representation conditions. The participants were required to generate a visual action representation in the Vi representation conditions and to generate an auditory verbal representation in the Ve representation conditions. Audition is considered a temporal sense, whereas vison is a spatial sense (Gori, et al., 2012; Kubovy, 1988), and simultaneous processing

is possible in vision (Millar, 1981). In the catch trials, simultaneous processing might occur during change detection in the Vi representation condition, whereas sequential processing might occur during change detection in the Ve representation condition. This modality difference might be related to the shorter reaction time seen in the Vi representation than in the Ve representation condition.

4.2. Transformation Process during the Stimulus Period

The fMRI results showed that the interaction effect of transformation ([ViVe + VeVi] > [ViVi + VeVe]) was significant in the left IFG (BA44/45). This activation was also found in the contrast of the verbal representation effect ([ViVe + VeVe] > [ViVi + VeVi]), and these two activation areas partially overlapped. The percent signal changes in the left IFG found in the interaction contrast may reflect the effect of verbal representation (i.e., an increase of the percent signal change in ViVe and VeVe), in addition to the interaction effect (i.e., an increase of the percent signal change in ViVe and VeVi). Although no statistical post-hoc test was performed on the percent signal changes, the percent signal change in VeVi seems to have increased more than in ViVi, and the percent signal change in ViVe seems to have increased more than in VeVe.

The left IFG is considered one of the key components of the mirror neuron system. The IFG and IPL are recruited by action execution and observation (Iacoboni & Dapretto, 2006; Rizzolatti et al., 2001; Rizzolatti & Sinigaglia, 2007, 2010). This indicates that these areas are involved in the process across action execution and visual action modalities. Some studies have suggested that the mirror neuron system is involved with the transformation between action execution information and visual action information (Gazzola et al., 2009; Iacoboni & Dapretto,

2006; Miall, 2003). There is a possibility that the left IFG activation found in the interaction contrast reflects that the left IFG stores amodal action representations and that the ViVe and VeVi routes (as shown in Fig. 1) need to pass through the area storing amodal action representations. However, a previous study with MVPA (Wurm & Caramazza, 2019) showed that common action representations across visual action and verbal modalities are stored in the lateral posterior temporal area, not in the front-parietal areas. Moreover, if the left IFG stores action representations across visual action and verbal modalities, then left IFG activation should also be observed in the ViVi and VeVe conditions, not just the ViVe and VeVi conditions. With this in mind, other studies are needed to interpret the results of the left IFG activation.

The prefrontal cortex, including the IFG, is considered involved with executive functions, such as working memory, inhibition, and attention control (Baddeley, 2010; Friedman & Robbins, 2022; Nee et al., 2013). Fedorenko et al. (2013) also reported that the frontal-parietal regions were engaged across various cognitive tasks, such as math, spatial working memory, and verbal working memory tasks. The results of the current study suggest that the ViVe and VeVi tasks required the cognitive processes related to executive functions more than the ViVi and VeVe tasks did; however, the significantly increased activity was found only in the left IFG, rather than in the various frontal-parietal regions. Therefore, it is important to discuss and identify the cognitive function of the left IFG regarding the transformation process required in the ViVe and VeVi tasks. We hypothesized that top-down or internally driven processes were more involved with generating internal representations in ViVe and VeVi than in ViVi and VeVe because the representation format types required by the tasks differed from the format types presented as stimuli in ViVe and VeVi, despite the required and presented format types being the same in ViVi and VeVe. Previous studies have shown that the IFG is involved with cognitive

and top-down processes, such as speech comprehension, speech syllable discrimination, letter detection, and face detection, under conditions where stimuli are ambiguous (Binder et al., 2004; Liu et al., 2010; Zekveld et al., 2006; Zhang et al., 2007). For example, Zekveld et al. (2006) reported that the increased activity in the left IFG (BA44) was found for unintelligible speech at low signal-to-noise ratios (in the low signal-to-noise ratio condition, their participants were presented with low intensity level spoken sentences masked with noise). They suggest that the left IFG is responsible for the top-down process to generate internal speech representations during the comprehending of unintelligible speech. A top-down signal from the frontal lobe is needed to generate mental imagery (Dijkstra et al., 2017; Ishai et al., 2000; Pearson, 2019). One interpretation of the left IFG activation found in the interaction contrast is that this region is related to a top-down process to generate a representation under a condition where an externally presented stimulus does not include information about a task's required representation itself.

Another interpretation of the left IFG activation is that, in addition to the cognitive process discussed above, other processes are involved in the ViVe and VeVi tasks. Some studies have reported that the IFG is one of the areas where activity increased during action imagery (Binkofski et al., 2000; Burianova et al., 2013; Ehrsson et al., 2003; Filimon et al., 2007; Hétu et al., 2013; Lotze & Halsband, 2006). It is possible that the left IFG activation in VeVi may include the involvement of the process of visual action imagery after the presentation of sentence stimuli. Word retrieval and binding words (syntactic unification) to make a sentence may be involved in ViVe. Retrieving words that correspond to a presented picture stimulus and generating a sentence by biding those words is required to perform the ViVe task appropriately. Word retrieval in the ViVe task seems to involve the selection of appropriate words among competing alternatives—in this case, the choice of appropriate words from *I* and *you*, *grasp* and

touch, and *apple* and *orange*. Previous studies have reported that the left IFG is involved in selecting information from competing alternatives (Badre et al., 2005; Heim et al., 2009; Moss et al., 2005; Robinson et al., 2010; Schnur et al., 2009; Thompson-Schill et al., 1997, 1998, 1999, 2005; Wagner et al., 2001). Hagoort (2003, 2013, 2014) and Hagoort and Indefrey (2014) postulated that the left IFG is related to syntactic binding (syntactic unification) and that BA45 and 44 are involved in syntactic processes, whereas BA 47 and 45 are involved in semantic processes and BA44 and the ventral BA6 are involved in phonological processes. The pars opercularis and pars triangularis in the left IFG (BA44/45) are known as Broca's area (Foundas et al., 1998; Toga & Thompson, 2003). Although it is involved in many functions (Ardila et al., 2016a, 2016b; Grodzinsky & Santi, 2008; Rogalsky & Hickok, 2011), the results of the present study suggest that at least one region of Broca's area was involved in transforming nonverbal picture stimuli describing actions into verbal representations, especially in tasks involving word selection among competitors and syntactic binding.

4.3. Verbal Representations during the Stimulus Period

In the stimulus period, a main effect of Ve representation ([ViVe + VeVe] > [ViVi + VeVi]) was found in the bilateral PreG and its surrounding areas (left IFG and bilateral MC, PMC, and PoG), left Spt, and bilateral preSMA, SMA, and cerebellum. Willems et al. (2010) suggested that the recognition of verbal stimuli describing action information involves implicit action simulation in the PMC, and different processes are involved with implicit action simulation and explicit action imagery. Therefore, it is possible that the PMC activation found in the main effect reflects implicit action simulation (or more abstract action simulation than specific visual action imagery) during the generation of the verbal representations.

Another interpretation of the main effect result is that it may reflect language processes during the generation of verbal representations. Although VeVe was the verbal format for both stimuli and representations, VeVe included the process of generating auditory (or speech) representations from the verbal letter stimuli. Therefore, it is assumed that both VeVe and ViVe were involved with the generation of auditory representations that were not included in the stimuli. Regarding the regions found in the main effect result, previous studies have reported their involvement in language processes, such as speech production, covert speech repetition, sensory-motor integration and transformation, and inner speech imagery (Amit et al., 2017; Ardila et al., 2016a; Frings et al., 2006; Hickok, 2012; Hickok et al., 2009; Hubrich-Ungureanu et al., 2002; Murdoch, 2010; Papathanassiou et al., 2000; Tian et al., 2016). In particular, we have focused on the bilateral PreG (including the PC) and the left IFG (Broca's area), in relation to action-related processes.

Both the PreG and Broca's area are thought to be involved in articulatory planning in speech production. Hickok (2012) proposed a model in which the areas including the PreG (ventral BA6 and MC) are involved in motor phoneme programs, and Broca's area (BA44) is involved in motor syllable programs. Although neuropsychological studies have traditionally focused on Broca's area as a key site for speech production, its role is controversial (Ardila, 2010; Ardila et al., 2016a, 2016b; Gajardo-Vidal et al., 2021; Lorca-Puls et al., 2021; Silva et al., 2022; Zhao et al., 2021). Studies using direct cortical electrical stimulation methods have revealed the importance of the ventral PreG for speech production (Lu et al., 2021; Tate et al., 2014; Zhao et al., 2021). In addition, a recent model of speech production postulated that the middle PreG is essential for speech motor planning (Silva et al., 2022). Neuropsychological studies (Basilakos et al., 2015; Itabashi et al., 2016) have reported that damage to the left PreG

is associated with apraxia of speech—a motor speech disorder in which impairment of the motor speech program is assumed. Studies have shown that articulatory planning is involved even in speech imagery without actual speech production and that it engages the frontal motor regions, including the IFG, PMC, SMA, and insula (Tian & Poeppel, 2012, 2013; Tian et al., 2016). According to these studies, it may be possible that the bilateral PreG (including the PMC) is involved in articulatory planning during the generation of verbal representations.

4.4. Internal Generation of Visual Action Representations in the Blank Period

During the blank period, the main effect of Vi representation ([ViVi + VeVi] > [ViVe + VeVe]) was found in the preSMA, left dPMC, and bilateral IPL. Previous studies have reported that these regions are engaged in the imagery of actions with an arm/hand effector (e.g., reaching actions) (Ehrsson, 2003; Filimon, 2007; Hétu, 2013). Research has also shown that the activities in motor-related areas during the recognition of action-related verbal stimuli are influenced by differences in tasks (Kulnke et al., 2020; 2021; van Dam et al., 2012; Willems et al., 2010; Yang & Shu, 2014). The regional activity found in this study might reflect an action imagery process with an arm/hand effector in the ViVi and VeVi tasks.

Although a Ve representation effect was found during the stimulus period, a Vi representation effect was observed during the blank period. One possible explanation for this difference is that the Vi representations were generated later than the Ve representations. If visual imagery is a similar experience to visual perception (Kosslyn et al., 2001; Miyashita, 1995), visual stimuli might interfere with generating visual representations when the visual stimuli are different from the visual representation—that is, it might be difficult to generate Vi representations during the observation of Ve stimuli. Although this interpretation may explain

the reason for the delay in the generation of Vi representations in the VeVi task, it cannot explain it in the ViVi task because the required representations were the same as for the presented stimuli. Another possible interpretation is that the activity in the areas showing the main effect of Vi representation (preSMA and MCC, left dPMC, and bilateral IPL) decreased during the blank period (or increased activity was not maintained during the blank period) in the ViVe and VeVe tasks. Previous studies have shown that both action-related verbal and nonverbal stimuli (e.g., video) increased activity in the regions including the PMC and the IPL, even when motor imagery of the presented stimuli was not required (Aziz-Zadeh et al., 2006; Filimon et al., 2007; Hauk et al., 2004), although some studies (Tomasino et al., 2007; 2014; Willems et al., 2010) have reported that different areas are involved in the understanding of verbal stimuli describing actions and the explicit action imagery of those stimuli, as mentioned above. It might be possible that no main effect of Vi representation was observed during the stimulus period because the activity in the areas including the preSMA, MCC, PMC, and IPL increased during the stimulus period in all four conditions (in which the action-related picture and sentence stimuli appeared), and the main effect of Vi representation was found during the blank period because the increased activity was maintained only in the ViVi and VeVi conditions during this period (in which the action-related picture and sentence stimuli disappeared and the retention of visual action representations was required).

4.5. Limitations of the Study and the Verbal and Visual Action Stimuli

First, we must acknowledge the limitation of interpreting the results found during the blank period. Because this began immediately after the end of the stimulus period, the brain activity observed during the blank period may have been influenced by the previously occurring

activity. The main effects of stimulus (Vi > Ve stimulus contrast and Ve > Vi stimulus contrast) during the stimulus period demonstrated activity change in the regions including the visual cortex in the occipital and temporal lobes. These results might reflect visual information processes for the Vi and Ve stimuli. It has been reported that some regions in the temporal and occipital lobes respond to visual stimuli, such as body parts and objects (Dehaene, Nakamura, et al., 2010; Dehaene, Pegado, et al., 2010; Dehaene & Cohen, 2011; Downing et al., 2001; Peelen & Downing, 2007). The activity we found in the temporal and occipital lobes regarding the main effect of Vi stimulus is thought to reflect this visual information process. However, the activity area found during the blank period for the Ve stimuli overlapped with the activity area observed in the stimulus period for the Vi stimuli (and the activity area during the blank period for the Vi stimuli overlapped with the activity area in the stimulus period for the Ve stimuli). It is unlikely that the process for the Vi stimuli in the Vi stimulus period (or for the Ve stimuli in the Ve stimulus period) was also found during the blank period after the Ve stimuli ended (or during the blank period after the Vi stimuli ended). Therefore, we speculate that the results found during the blank period for the Vi and Ve stimuli cannot exclude the effect of activity changes that occurred during the stimulus period. Similarly, the results of the main effect of Vi representation found during the blank period may be affected by the activity that occurred during the stimulus period. This is a problem with the experimental design, thus the interpretation of the results found during the blank period is a noteworthy limitation.

Another limitation is the interpretation of left IFG activation found in the interaction contrast. We interpreted the results by using a subtraction method on the whole-brain analysis. Although the conjunction analysis of ViVe transformation contrast ([ViVe > ViVi] ∩ [ViVe > VeVe]) did not show a significant difference, 10 voxels survived in the left IFG with a threshold

of $p < 0.001$, uncorrected for multiple comparisons at the voxel level. Therefore, it might be possible that the left IFG activation was more strongly influenced by ViVe than by VeVi. It might also be possible that different processes are involved under the ViVe and VeVi tasks. Additional studies with additional analyses, such as an ROI analysis and an MVPA for the left IFG area found in this study, would be helpful for the further clarification of the cognitive processes related to the ViVe and VeVi tasks.

5. Conclusion

We conducted an fMRI study in which the format types of input stimulus and internal output representation were independently controlled to clarify the transformation process of the format types between the stimuli and representations. The results showed that the left IFG (BA44/45) is involved in the transformation process. We discussed the function of the left IFG and suggested that this area was involved with generating a representation under a condition where externally presented stimulus did not include information concerning the task's required representation itself.

Authorship Contributions

Hiroshi Shibata: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing - original draft, Writing - review & editing.

Kenji Ogawa: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing - review & editing.

Funding

Shibata and Ogawa were supported in part by a grant from the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Number JP18K12019).

Conflicts of Interest

We have no conflicts of interest to declare.

Acknowledgments

We are sincerely grateful to the anonymous reviewers for their comments, which helped improve the manuscript.

References

Amit, E., Hoeflin, E., Hamzah, N., & Fedorenko, E. (2017). An asymmetrical relationship between verbal and visual thinking: converging evidence from behavior and fMRI. *NeuroImage, 152*, 619-627.

Ardila, A. (2010). A proposed reinterpretation and reclassification of aphasic syndromes, *Aphasiology*, *24*(3), 363-394.

Ardila, A., Bernal, B., & Rosselli, M. (2016a). How localized are language brain areas? A review of Brodmann areas involvement in oral language. *Archives of Clinical Neuropsychology*, *31*, 112–122.

Ardila, A., Bernal, B., & Rosselli, M. (2016b). Why Broca's area damage does not result in classical Broca's aphasia. *Frontiers in Human Neuroscience*, *10*, 249. https://doi: 10.3389/fnhum.2016.00249.

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, *16*(18), 1818–1823.

Baddeley, A. (2010). Working memory. Current Biology, 20(4), R136-140.

Badre, D., Poldrack, R.A., Paré-Blagoev, E.J., Insler, R.Z., & Wagner, A.D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*, 907-918.

Barsalou, L.W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.

Basilakos, A., Rorden, C., Bonilha, L., Moser, D., & Fridriksson, J. (2015). Patterns of poststroke brain damage that predict speech production errors in apraxia of speech and

aphasia dissociate. *Stroke*, *46*(6), 1561–1566.

Binder, J., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience, 7*(3), 295-301.

Binkofski, F., Amunts, K., Stephan, K. M., Posse, S., Schormann, T., Freund, H. J., Zilles, K., & Seitz, R. J. (2000). Broca's region subserves imagery of motion: a combined cytoarchitectonic and fMRI study. *Human Brain Mappipng, 11*, 273–285.

Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cerebral Cortex, 19*(8), 1905-1914.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freud, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, *13*(2), 400–404.

Burianová, H., Marstaller, L., Sowman, P., Tesan, G., Rich, A.N., Williams, M., Savage, G., & Johnson, B.W. (2013). Multimodal functional imaging of motor imagery using a novel paradigm. *NeuroImage, 71*, 50-58.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., Mazziotta, J. C., & Fazio, F. (1994). Mapping motor representations with positron emission tomography. *Nature*, *371*, 600-602.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254–262.

Dehaene, S., Nakamura, K., Jobert, A., Kuroki, C., Ogawa, S., & Cohen, L. (2010). Why do

children make mirror errors in reading? Neural correlates of mirror invariance in the visual word form area. *NeuroImage*, *49*(2), 1837–1848.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*(6009), 1359–1364.

Dijkstra, N., Zeidman, P., Ondobaka, S. van Gerven, M. A. J., & Friston, K. (2017). Distinct top-down and bottom-up brain connectivity during visual perception and imagery, *Scientific Reports, 7*:5677. https://doi.org/10.1038/s41598-017-05888-8.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473.

Ehrsson, H. H., Geyer, S., & Naito, E. (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. *Journal of Neurophysiology*, *90*, 3304-3316.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *The proceedings of the National Academy of Sciences*, *110*(41), 16616-16621.

Filimon, F., Nelson, J. D., Hagler, D. J., & Sereno, M. I. (2007). Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *NeuroImage*, *37*, 1315-1328

Foundas, A. L., Eure, K. F., Luevano, L. F., & Weinberger, D. R. (1998). MRI asymmetries of Broca's area: the pars triangularis and pars opercularis. *Brain and Language*, *64*(3), 282-296.

Friedman, N. P., & Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology, 47*, 72-89.

Frings, M., Dimitrova, A., Schorn, C. F., Elles, H.-G., Hein-Kropp, C., Gizewski, E. R., Diener, H. C., & Timmann, D. (2006). Cerebellar involvement in verb generation: an fMRI study. *Neuroscience Letters*, *409*(1), 19-23.

Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex, 19*, 1239-1255.

Gajardo-Vidal, A., Lorca-Puls, D. L., Team P, Warner, H., Pshdary, B. Crinion, J. T., Leff, A. P., Hope, T. M. H., Geva, S., Seghier, M. L., Green, D. W., Bowman, H., & Price, C. J. (2021). Damage to Broca's area does not contribute to long-term speech production outcome after stroke. *Brain*, *144*(3), 817–832.

Gori M., Sandini, G., & Burr, D. (2012). Development of visuo-auditory integration in space and time. *Frontiers in Integrative Neuroscience, 6*(77). https://doi.org/10.3389/fnint.2012.00077.

Grodzinsky, Y. & Santi, A. (2008). The battle for Broca's region. *Trends in Cognitive Sciences*, *12*(12), 474-480.

Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *NeuroImage*, *20*, S18-S29.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00416.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology*, *28*, 136-141.

Hagoort, P. & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, *37*, 347-362.

Hanakawa, T. (2016). Organizing motor imageries. *Neuroscience Research*, *104*, 56-63.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.

Heim, S., Eickhoff, S.B., Friederici, A. D., & Amunts, K. (2009). Left cytoarchitectonic area 44 supports selection in the mental lexicon during language production. *Brain Structure and Function*, *213*, 441-456.

Hétu, S., Grégoire, M. Saimpont, A., Coll, M.-P., Eugéne, F., Michon, P.-E., & Jackson, P. L. (2013). The neural network of motor imagery: an ALE meta-analysis. *Neuroscience and Biobehavioral Reviews*, *37*(5), 930-949.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145.

Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, *101*(5), 2725–2732.

Hubrich-Ungureanu, P., Kaemmerer, N., Henn, F. A., & Braus, D. F. (2002). Lateralized organization of the cerebellum in a silent verbal fluency task: a functional magnetic resonance imaging study in healthy volunteers. *Neuroscience Letters*, *319*(2), 91-94.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neurosciences*, *7*, 942-951.

Ishai, A., Ungerleider, L.G., & Haxby, J.V. (2000). Distributed neural systems for the generation

of visual images. *Neuron, 28*, 979-990.

Itabashi, R., Nishio, Y., Kataoka, Y., Yazawa, Y., Furui, E., Matsuda, M., & Mori, E. (2016). Damage to the left precentral gyrus is associated with apraxia of speech in acute stroke. *Stroke*, *47*(1), 31–36.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*. *2*(9), 635-642.

Kubovy, M. (1988). Should we resist the seductiveness of the space: time:: vision: audition analogy? *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 318-320.

Kuhnke, P., Kiefer, M., & Hartwigsen, G. (2020). Task-dependent recruitment of modality-specific and multimodal regions during conceptual processing. *Cerebral Cortex, 30*, 3938-3959.

Kuhnke, P., Kiefer, M., & Hartwigsen, G. (2021). Task-dependent functional and effective connectivity during conceptual processing. *Cerebral Cortex, 31*, 3475-3493.

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2016). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*, 42–55.

Lorca-Puls, D. L., Gajardo-Vidal, A., Green, D. W., & Price, C. J. (2021). Reply: Broca's area: why was neurosurgery neglected for so long when seeking to re-establish the scientific truth? and Where is the speech production area? Evidence from direct cortical electrical stimulation mapping. *Brain*, *144*(7), e62, https://doi.org/10.1093/brain/awab177.

Lotze, M., & Halsband, U. (2006). Motor imagery. *Journal of Physiology - Paris, 99*, 386-395.

Liu J., Li, J., Zhang, H., Rieth, C. A., Huber, D. E., Li, W., Lee, K., & Tian, J. (2010). Neural

correlates of top-down letter processing. *Neuropsychologia, 48*, 636-641.

Lu, J., Zhao, Z., Zhang, J., Wu, B., Zhu, Y., Chang, E. F., Wu, J., Duffau, H., & Berger, M. S. (2021). Functional maps of direct electrical stimulation-induced speech arrest and anomia: a multicentre retrospective study. *Brain*, *144*(8), 2541-2553.

Martin A., Haxby, J.V., Lalonde, F.M., Wiggs, C.L., & Ungerleider, L.G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science, 270*, 102-105.

Miall, R.C. (2003). Connecting mirror neurons and forward models. *NeuroReport, 14*(17), 2135-2137.

Millar, S. (1981). Crossmodal and intersensory perception and the blind. In R.D. Walk, & H.L. Pick, JR (Eds.) *Intersensory Perception and Sensory Integration. Perception and Perceptual Development* (pp.281-314). Springer.

Miyashita, Y. (1995). How the brain creates imagery: projection to primary visual cortex. Science, *268*(5218), 1719-1720.

Moss, H. E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., & Tyler, L. K. (2005). Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus. *Cerebral Cortex*, *15*(11), 1723-1735.

Murdoch, B. E. (2010). The cerebellum and language: historical perspective and review. *Cortex, 46*, 858-868.

Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., Jonides, J. (2013). A meta-analysis of executive components of working memory. *Cerebral Cortex, 23*(2), 264-282.

Nicholls, M. E. R., Thomas, N. A., Loetscher, T., & Grimshaw, G. M. (2013). The flinders handedness survey (FLANDERS): a brief measure of skilled hand preference. *Cortex*, *49*, 2914–2926.

Okubo, M., Suzuki, H., & Nicholls, M.E.R. (2014). A Japanese version of the FLANDERS handedness questionnaire. *Shinrigaku Kenkyu: The Japanese Journal of Psychology*, *85*, 474–481.

Papathanassiou, D., Etard, D., Mellet, E., Zago, L., Mazoyer, B., & Tzourio-Mazoyer, N. (2000). A common language network for comprehension and production: a contribution to the definition of language epicenters with PET. *NeuroImage*, *11*(4), 347-357.

Patterson, K., Nestor, P.J., & Rogers, T.T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976-988.

Pavio, A. (2013). Mind and its evolution: a dual coding theoretical approach. New York: Psychology Press.

Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience, 20*, 624-634.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*(8), 636–648.

Price, A. R., Bonner, M. F., & Grossman, M. (2015). Semantic memory: cognitive and neuroanatomical perspectives. *Brain Mapping: An Encyclopedic Reference* (Vol. 3). Elsevier Inc.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of

heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847.

Price, C. J. (2018). The evolution of cognitive models: from neuropsychology to neuroimaging and back. *Cortex*, *107*, 37–49.

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388–396.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*, 661-670.

Rizzolatti, G., & Sinigaglia, C. (2007). Mirror neurons and motor intentionality. *Functional Neurology*, *22*(4), 205-210.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*(4), 264-274.

Robinson, G., Shallice, T., Bozzali, M., & Cipolotti, L. (2010). Conceptual proposition selection and the LIFG: neuropsychological evidence from a focal frontal group. *Neuropsychologia*, *48*(6), 1652-1663.

Rogalsky, C. & Hickok, G. (2011). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience*, *23*(7), 1664-1680.

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences*, *106*(1), 322–327.

Shibata, H., & Ogawa, K. (2018). Dorsal premotor cortex is related to recognition of verbal and

visual descriptions of actions in the first-person perspective. *Neuroscience Letters*, *687*, 71–76.

Silva, A. B., Liu, J. R., Zhao, L., Levy, D. F., Scott, T. L., & Chang, E. F. (2022). A neurosurgical functional dissection of the middle precentral gyrus during speech production. *Journal of Neuroscience*, *42*(45), 8416-8426.

Tate, M. C., Herbet, G., Moritz-Gasser, S., Tate, J. E., & Duffau, H. (2014). Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited. *Brain*, *137*, 2773-2782.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences*, *94*, 14792-14797.

Thompson-Schill, S. L., Swick, D., Farah, M. J., D'Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: a neuropsychological test of neuroimaging findings. *Proceedings of the National Academy of Sciences*, *95*, 15855-15860.

Thompson-Schill, S. L., D'Esposito, M., & Kan, I. P. (1999). Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron*, *23*, 513-522.

Thompson-Schill, S. L., Bedny, M., Goldgerg, R. F. (2005). The frontal lobes and the regulation

of mental activity. *Current Opinion in Neurobiology*, *15*, 219-224.

Tian X., & Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, *6*, 314. https://doi: 10.3389/fnhum.2012.00314.

Tian X., & Poeppel, D. (2013). The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience*, *25*(7), 1020-1036.

Tian X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*, *77*, 1-12.

Toga, A. W., & Thompson, P. M. (2003). Mapping brain asymmetry. *Nature Reviews Neuroscience*, *4*, 37-48.

Tomasino, B., Werner, C. J., Weiss, P. H., & Fink, G. R. (2007). Stimulus properties matter more than perspective: an fMRI study of mental imagery and silent reading of action phrases. *NeuroImage, 36*, T128-T141.

Tomasino, B., & Rumiati, R. I. (2013). At the mercy of strategies: the role of motor representations in language understanding. *Frontiers in Psychology, 4*:27. https://doi.org/10.3389/fpsyg.2013.00027.

Tomasino, B., Fabbro, F., & Brambilla, P. (2014). How do conceptual representations interact with processing demands: an fMRI study on action- and abstract-related words. *Brain Research, 1591*, 38-52.

van Dam, W.O., van Dijk, M., Bekkering, H., & Rueschemeyer, S-A. (2012). Flexibility in embodied lexical-semantic representations. *Human Brain Mapping, 33*, 2322-2333.

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*, 329-338.

Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2010). Neural dissociations between action verb understanding and motor imagery. *Journal of Cognitive Neuroscience, 22*(10), 2387-2400.

Wurm, M. F., & Caramazza, A. (2019). Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language, *Nature Communications*, *10*:289. https://doi.org/10.1038/s41467-018-08084-y.

Wurm, M. F., & Caramazza, A. (2022). Two 'what' pathways for action and object recognition, *Trends in Cognitive Sciences*, *26*(2), 103-116.

Yang, J., & Shu, H. (2014). Passive reading and motor imagery about hand actions and tool-use actions: an fMRI study. *Experimental Brain Research*, *232*(2), 453–467.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage, 32*, 1826-1836.

Zhang, H., Liu, J., Huber, D. E., Rieth, C. A., Tian, J., & Lee, K. (2007). Detecting faces in pure noise images: a functional MRI study on top-down perception. *NeuroReport, 19*(2), 229-233.

Zhao, Z., Liu, Y., Zhang, J., Lu, J., & Wu, J. (2021). Where is the speech production area? Evidence from direct cortical electrical stimulation mapping. *Brain*, *144*(7), e61. https://doi.org/10.1093/brain/awab178.

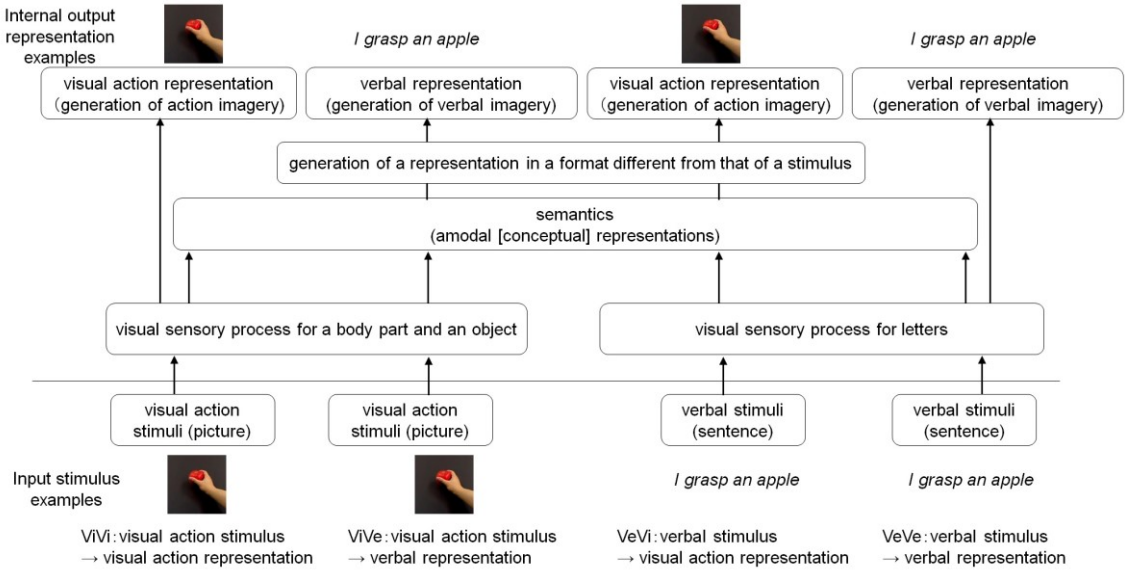


Fig. 1. Outline of the cognitive processes involved in generating visual action and verbal representations from visual action and verbal stimuli.

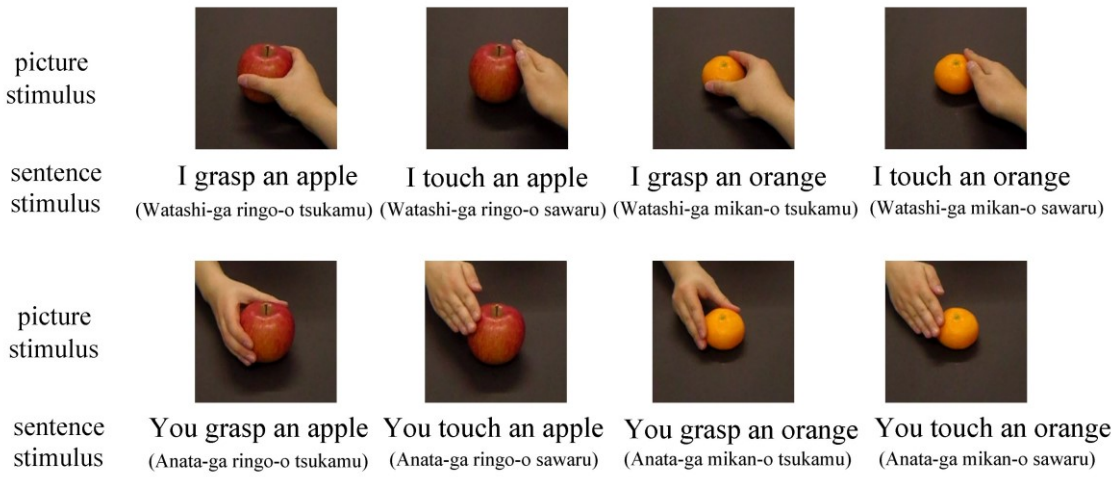
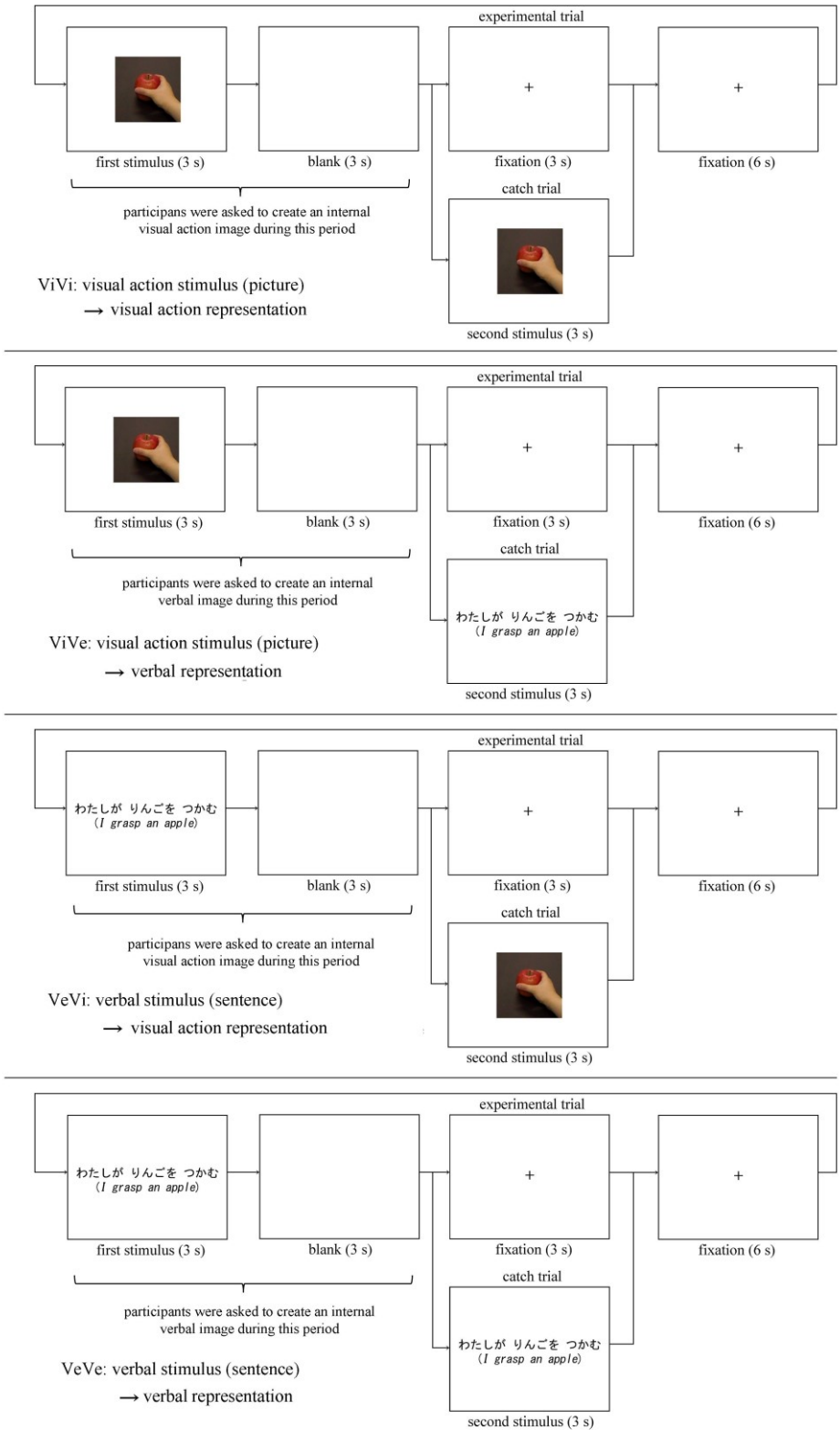


Fig. 2. Eight types of visual action stimuli (pictures) and verbal stimuli (sentences) used in this study. Verbal stimuli were presented in Japanese in the experiment.



1

2 Fig. 3. Schematic of the experimental tasks in ViVi, ViVe, VeVi, and VeVe.

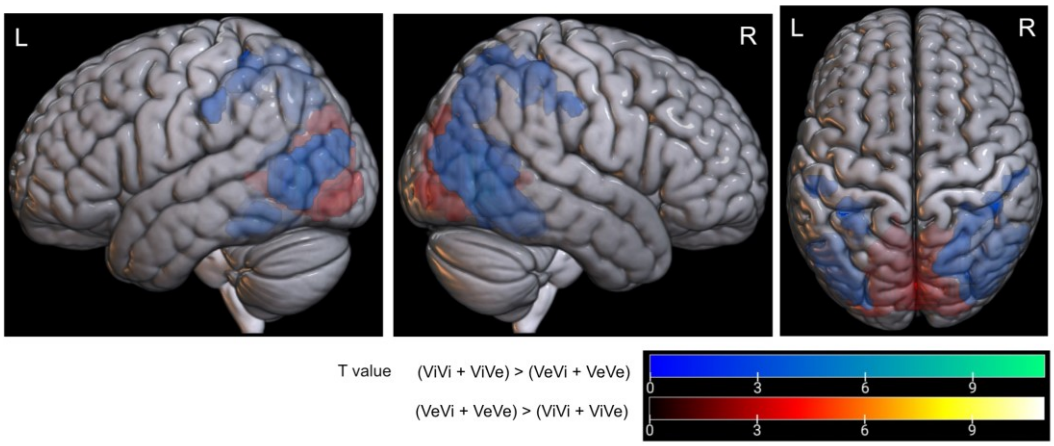


Fig. 4. Brain regions found in the main effects of stimulus in the stimulus period. The blue regions show a significantly greater activity in the visual action stimulus condition than in the verbal stimulus condition ([ViVi + ViVe] > [VeVi + VeVe]). The red regions show a significantly greater activity in the verbal stimulus condition than in the visual action stimulus condition ([VeVi + VeVe] > [ViVi + ViVe]).

L: left; R: right.

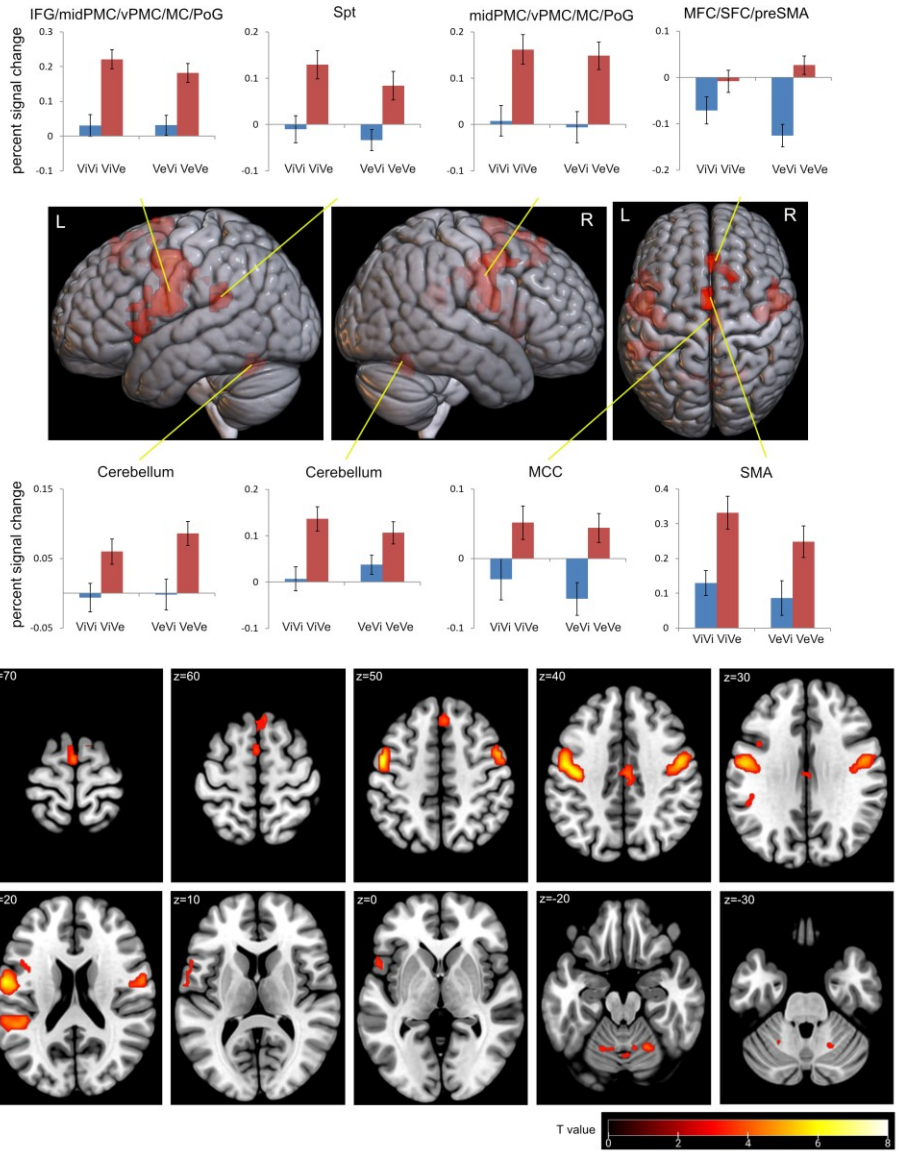


Fig. 5. Brain regions found in the main effect of representation ([ViVe + VeVe] > [ViVi + VeVi]) in the stimulus period and the mean percent signal changes in each region.

IFG: inferior frontal gyrus; midPMC: mid-premotor cortex; vPMC: ventral premotor cortex; MC: primary motor cortex; PoG: postcentral gyrus; Spt: Sylvian-parietal-temporal region; MFC: medial frontal cortex; SFC: superior frontal cortex; preSMA: pre-supplementary motor area; MCC: middle cingulate cortex; SMA: supplementary motor area; L: left; R: right.

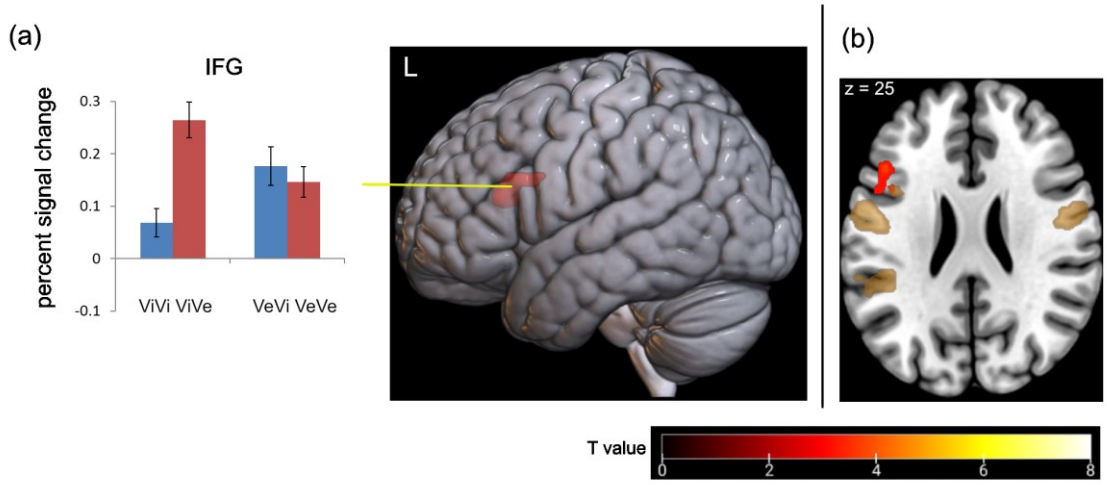


Fig. 6. (a) Brain region found in the interaction effect ([ViVe + VeVi] > [ViVi + VeVe]) in the stimulus period, and the mean percent signal changes in the region. (b) Brain regions found in the interaction effect ([ViVe + VeVi] > [ViVi + VeVe]) (red region) and the main effect of representation ([ViVe + VeVe] > [ViVi + VeVi]) (gold regions).

IFG: inferior frontal gyrus; L: left.

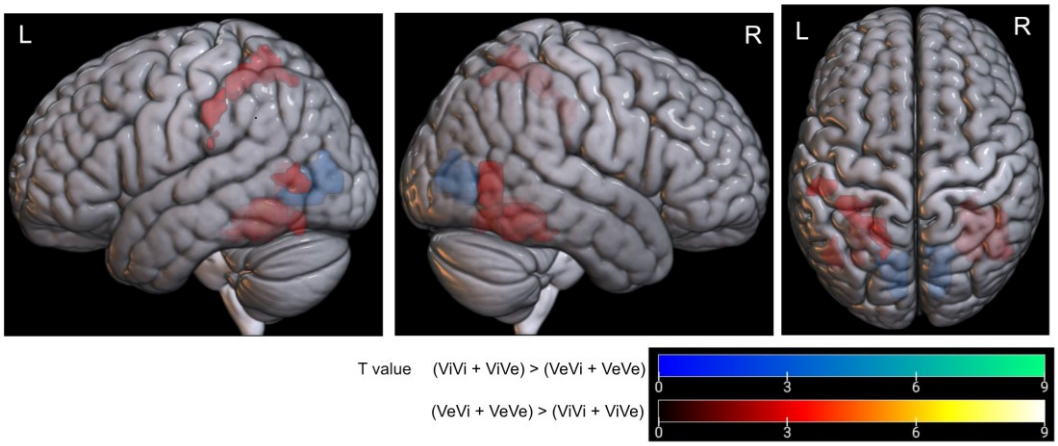


Fig.7. Brain regions found in the main effects of stimulus in the blank period. The blue regions show a significantly greater activity in the visual action stimulus condition than in the verbal stimulus condition ([ViVi + ViVe] > [VeVi + VeVe]). The red regions show a significantly greater activity in the verbal stimulus condition than in the visual action stimulus condition ([VeVi + VeVe] > [ViVi + ViVe]).

L: left; R: right.

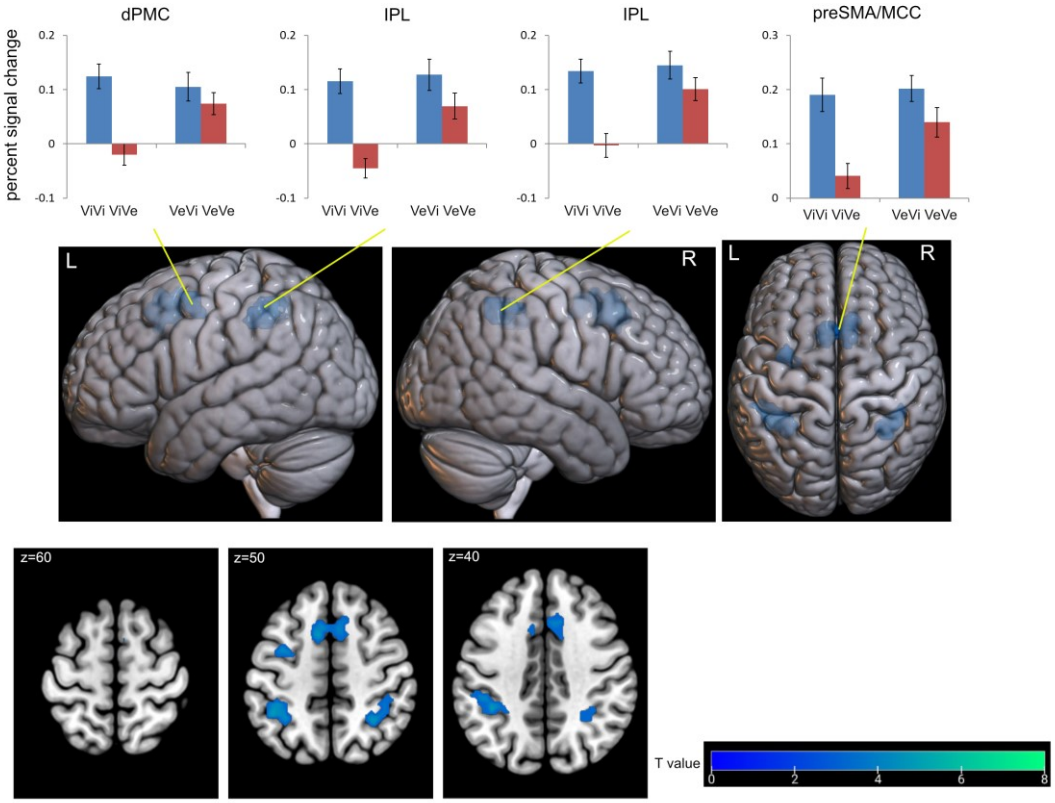


Fig. 8. Brain regions found in the main effect of representation ([ViVi + VeVi] > [ViVe + VeVe]) in the blank period and the mean percent signal changes in each region.

dPMC: dorsal premotor cortex; IPL: inferior parietal lobule; preSMA: pre-supplementary motor area; MCC: middle cingulate cortex; L: left; R: right.

Table 1

Brain regions related with the stimulus period.

Anatomic region	BA	MNI Coordinates X	Y	Z	Voxels	*t* value
1. Main effect of stimulus: Vi > Ve stimulus ([ViVi + ViVe] > [VeVi + VeVe])						
L inferior parietal lobule/supramarginal gyrus	2/3	-51	-28	38	58	5.40
L inferior parietal lobule/ superior parietal lobule	2/7/40	-33	-46	59	91	5.00
L postcentral gyrus	2	-39	-43	65		3.78
L fusiform gyrus/cerebellum	37	-30	-49	-16	67	5.32
L fusiform gyrus	37	-30	-37	-22		5.02
R inferior temporal gyrus/middle temporal gyrus extending into superior occipital gyrus and inferior and parietal lobes	2/3/7/20 /37/39/40	48	-61	-4	1493	10.48
R fusiform gyrus/cerebellum	37	33	-49	-16		7.73
R middle temporal gyrus/middle occipital gyrus	18/19/37	48	-70	14		7.42
L middle temporal gyrus/inferior temporal gyrus	37	-48	-67	2	522	8.36
L middle occipital gyrus/inferior occipital gyrus	18/19	-33	-82	11		6.44
L superior occipital gyrus	18/39	-27	-91	23		4.91
2. Main effect of stimulus: Ve > Vi stimulus ([VeVi + VeVe] > [ViVi + ViVe])						
L lingual gyrus/ calcarine cortex/ inferior occipital gyrus/cuneus	17/18/19	-9	-88	-1	1932	10.00
R lingual gyrus/calcarine cortex/inferior occipital gyrus/cuneus	17/18/19	12	-85	2		9.70

R calcarine cortex	17	15	-82	11		8.99
3. Main effect of representation: Ve > Vi representation ([ViVe + VeVe] > [ViVi + VeVi])						
medial frontal cortex/pre-supplementary motor area	8	0	29	56	88	4.31
R superior frontal cortex/pre-supplementary motor area	32	12	23	47		4.07
R superior frontal cortex/pre-supplementary motor area	6	18	17	65		3.99
L supplementary motor area	6	-3	-1	74	58	5.10
supplementary motor area	6	0	5	59		3.80
R mid- premotor cortex	6	51	-4	44	288	6.10
R postcentral gyrus	3	42	-10	35		5.50
R ventral premotor cortex/primary motor cortex	4	51	-4	26		5.39
L mid- premotor cortex/primary motor cortex/postcentral gyrus	4/6	-51	-7	47	493	7.54
L ventral premotor cortex/inferior frontal gyrus (opercularis/triangularis)/postcentral gyrus	43/44/45	-60	-4	20		6.46
L postcentral gyrus	3	-42	-16	41		6.28
R middle cingulate cortex	23	3	-16	38	48	4.79
L middle cingulate cortex	23	-3	-16	29		3.50
L middle cingulate cortex	23	-3	-31	35		3.41
L sylvian-parietal-temporal area	41/42	-51	-40	20	127	4.76
L sylvian-parietal-temporal area		-60	-37	23		4.62
L cerebellum	18/19	-21	-61	-22	48	4.48
R cerebellum		3	-67	-22		4.20

R cerebellum	18	9	-58	-19		3.87
R cerebellum	19/37	21	-61	-25	40	5.90
4. Interaction: transformation effect ([ViVe + VeVi] > [ViVi + VeVe])						
L inferior frontal gyrus (triangularis)	45	-42	26	17	54	4.30
L inferior frontal gyrus (opercularis)	44	-42	8	26		3.48
L inferior frontal gyrus (opercularis)		-48	14	23		3.44

Results thresholds comprise height of $p < 0.001$ uncorrected for multiple comparisons and $p <$ 0.05 FWE-corrected for the cluster level. Up to three local maxima more than 8 mm apart within each cluster are shown.

BA = Brodmann area; MNI = Montreal Neurological Institute; L = Left; R = Right.

Table 2

Brain regions related with the blank period.

Anatomic region	BA	MNI Coordinates X	Y	Z	Voxels	*t* value
1. Main effect of stimulus: Vi > Ve stimulus ([ViVi + ViVe] > [VeVi + VeVe])						
R lingual gyrus/ calcarine cortex	17/18	18	-76	2	129	6.20
R lingual gyrus	18	6	-64	5		3.87
L lingual gyrus/ calcarine cortex	17	-9	-85	2	161	5.75
L lingual gyrus	18	-12	-70	-1		5.63
2. Main effect of stimulus: Ve > Vi stimulus ([VeVi + VeVe] > [ViVi + ViVe])						
L inferior parietal lobule	2	-51	-28	38	51	5.04
L supramarginal gyrus		-54	-25	20		3.99
L inferior parietal lobule/ postcentral gyrus	2/3/40	-39	-37	53	154	5.27
L inferior parietal lobule/ superior parietal lobule	5/7	-18	-55	56		4.71
L inferior parietal lobule/ superior parietal lobule	2/7/40	-30	-46	59		4.50
L fusiform gyrus	37	-33	-52	-13	135	8.07
L fusiform gyrus/ cerebellum	37	-27	-37	-22		4.46
L fusiform gyrus	19	-33	-67	-16		4.24
R fusiform gyrus	19/37	36	-55	-16	344	7.48
R fusiform gyrus/cerebellum	37	30	-40	-19		6.81
R inferior temporal gyrus	20	48	-46	-16		4.98
L middle temporal gyrus/middle occipital gyrus	37	-48	-67	5	42	4.59

L middle temporal gyrus/middle occipital gyrus	37	-45	-58	5		3.30
3. Main effect of representation: Vi > Ve representation ([ViVi + VeVi] > [ViVe + VeVe])						
R middle cingulate cortex/pre-supplementary motor area	6/32	9	20	44	193	5.13
L middle cingulate cortex/pre-supplementary motor area	32	-9	11	47		5.13
L pre-supplementary motor area	6	-6	17	53		4.63
L dorsal premotor cortex	6	-30	-4	44	44	4.59
L dorsal premotor cortex	6	-33	-1	53		4.48
L inferior parietal lobule	40	-42	-46	47	150	6.18
L inferior parietal lobule	40	-33	-49	50		4.40
R inferior parietal lobule	7/40	30	-49	44	93	4.89
R inferior parietal lobule	40	39	-40	47		3.68

Results thresholds comprise height of $p < 0.001$ uncorrected for multiple comparisons and $p <$ 0.05 FWE-corrected for the cluster level. Up to three local maxima more than 8 mm apart within each cluster are shown.

BA = Brodmann area; MNI = Montreal Neurological Institute; L = Left; R = Right.