

Cross-modal Binding and Activated Attentional Networks during Audio-visual Speech Integration: a Functional MRI Study

Daisuke N. Saito¹, Kumiko Yoshimura¹, Takanori Kochiyama³, Tomohisa Okada¹, Manabu Honda¹ and Norihiro Sadato^{1,2}

¹National Institute for Physiological Sciences, Okazaki, Japan, ²JST (Japan Science and Technology Corporation)/RISTEX (Research Institute of Science and Technology for Society), Kawaguchi, Japan and ³Faculty of Human Studies, Kyoto University, Kyoto, Japan

We evaluated the neural substrates of cross-modal binding and divided attention during audio-visual speech integration using functional magnetic resonance imaging. The subjects ($n = 17$) were exposed to phonemically concordant or discordant auditory and visual speech stimuli. Three different matching tasks were performed: auditory-auditory (AA), visual-visual (VV) and auditory-visual (AV). Subjects were asked whether the prompted pair were congruent or not. We defined the neural substrates for the within-modal matching tasks by VV-AA and AA-VV. We defined the cross-modal area as the intersection of the loci defined by AV-AA and AV-VV. The auditory task activated the bilateral anterior superior temporal gyrus and superior temporal sulcus, the left planum temporale and left lingual gyrus. The visual task activated the bilateral middle and inferior frontal gyrus, right occipito-temporal junction, intraparietal sulcus and left cerebellum. The bilateral dorsal premotor cortex, posterior parietal cortex (including the bilateral superior parietal lobule and the left intraparietal sulcus) and right cerebellum showed more prominent activation during AV compared with AA and VV. Within these areas, the posterior parietal cortex showed more activation during concordant than discordant stimuli, and hence was related to cross-modal binding. Our results indicate a close relationship between cross-modal attentional control and cross-modal binding during speech reading.

Keywords: cross-modal matching, human voice, integration, intraparietal sulcus, visual

Introduction

The integration of information conveyed through anatomically distinct sensory pathways is necessary for many human behaviors. In particular, auditory and visual multimodal information processing is important for face-to-face communication. Multi-sensory convergence contributes to the interpretation of the functional significance of stimuli linked by a common causality (Meredith *et al.* 1987). Speech reading, or lipreading, for example, enhances the perception of spoken language. Combining audible speech with the corresponding visible articulation movements can improve comprehension to the same degree as altering the acoustic signal-to-noise ratio by 15–20 dB (Sumbly and Polack, 1954). Several functional neuroimaging approaches have been used to study cross-modal integration in humans. The first indication of audio-visual integration data emerged when activity in the auditory areas was seen during the viewing of visual stimuli. Calvert *et al.* (1997) found that silent lipreading activated the superior temporal gyrus. Calvert *et al.* (2000) also found that the left superior temporal sulcus (STS) exhibited a significant response enhancement to congruent audio-visual inputs; this enhancement was supra-additive to that seen in

each isolated modality. Hence, they suggested that STS is one site of audio-visual integration.

However, several studies report conflicting results. Using temporally synchronized and desynchronized voice and lip movement stimuli, Olson *et al.* (2002) found no activation in the STS; rather, they found activation in the claustrum. Jones and Callan (2003) used functional magnetic resonance imaging (fMRI) to assess the relationship between brain activity and the degree of audio-visual integration of speech information during a phoneme categorization task. They found that visual information had a strong influence on speech perception (the McGurk effect), and this was positively correlated with the activity in the left occipito-temporal junction, an area often associated with processing visual motion. Based on these results, they proposed that auditory information modulates visual processing to affect perception. Bushara *et al.* (2003) measured transient brain responses to audio-visual binding, seen during a sound-induced change in visual motion perception. They found that cross-modal binding was associated with higher activity in multimodal areas, including the insula/frontal operculum, dorsolateral and medial prefrontal cortex, posterior parietal cortex, posterior thalamus, superior colliculus and posterior cerebellar vermis. Using audio-visual motion discrimination tasks, Lewis *et al.* (2000) found activation in the intraparietal sulcus (IPS) and fronto-parietal network, suggesting a close relationship between attentional selection and cross-modal integration. They did not find cross-modal activation in the STS. In summary, the proposed neural substrates of audio-visual cross-modal integration, particularly between voice sounds and lip movements, remain controversial.

The comparison between cross-modal and unimodal processing should take into account attentional modulation, which may in part explain the discrepancy across studies (Lewis *et al.*, 2000). Subjects must attend to both modalities during cross-modal conditions, but only to a single modality during unimodal conditions. We hypothesized that an auditory and visual cross-modal speech matching task should involve both divided attention and cross-modal integration processes that are not required for the unimodal matching tasks. Hence, comparisons between the cross-modal and unimodal conditions would highlight the neural substrates underlying divided attention and cross-modal binding. As behavioral results have revealed cross-modal facilitation during lipreading (Sumbly and Polack, 1954), the neural substrates of the mechanisms underlying cross-modal binding can be evaluated based on the modulation of responses due to the congruency/incongruency of the stimuli.

In the present study, we explicitly controlled the attentional modulation in the voice and lip movement matching tasks by preparing audio-visual stimuli in which subjects attend only to

the appropriate modality. Stimuli are identical across the conditions, but the subject is instructed either to direct their attention to one or the other modality or divide their attention between the modalities. A direct comparison between cross-modal and unimodal tasks was planned in order to clarify the net effect of both attentional modulation and cross-modal binding. Within the areas showing the sum of these effects, we compared the congruent and incongruent cross-modal conditions, subtracting out the effects of attention.

Materials and Methods

Subjects

Seventeen healthy volunteers (10 men and 7 women, mean age 27.8 ± 6.7 years) participated in this study. Sixteen subjects were right-handed and one was left-handed according to the Edinburgh handedness inventory (Oldfield, 1971). None of the subjects had a history of neurological or psychiatric illness. The protocol was approved by the ethical committee of the National Institute for Physiological Sciences, and all subjects gave their written informed consent for the study.

MRI

A time-course series of 124 volumes were acquired using T2*-weighted, gradient echo, echo planar imaging (EPI) sequences using a 3.0 T MR imager (Allegra, SIEMENS, Erlangen, Germany). Each volume consisted of 36 slices, each 3.0 mm thick, with a 0.6 mm gap, to cover the entire cerebral and cerebellar cortex. Oblique scanning was used to exclude the eyeballs from the images. The time interval between two successive acquisitions of the same slice was 4000 ms with a flip angle of 85 degrees and 30 ms echo time. The cluster volume acquisition time was 2400 ms, leaving a 1600 ms silent period (Edmister *et al.*, 1999). The field of view (FOV) was 192 mm and the in-plane matrix size was 64×64 pixels. For anatomical reference, T1-weighted MPRAGE [$T_R = 1460$ ms, $T_E = 4.38$ ms, field angle (FA) = 8° , FOV = 192 mm, matrix size = 256×256 mm] were collected at the same positions as the echo planar images, and 3-D MPRAGE ($T_R = 2500$ ms, $T_E = 4.38$ ms, FA = 8° , FOV = 230 mm, matrix size = 256×256 mm, slice thickness = 1 mm, a total of 192 transaxial images) were obtained for each subject.

Face-Voice Matching Tasks

For the face-voice matching tasks, the stimuli were of a human face pronouncing vowels. The auditory and visual stimuli consisted of a digitally recorded female voice and face (16 bit, 11.025 kHz sampling rate, using Adobe Premiere, Adobe, San Jose, CA). The maximum sound pressure (72 dB at the ear), frequency range and duration of each stimuli were adjusted and presented via earphones using Presentation software (Neurobehavioral Systems, Albany, CA) on a microcomputer (Dimension 8200, Dell Computer Co., Round Rock, TX). To minimize the effects of the noise of the MRI scanner, the auditory stimuli were presented during an interval of scanner silence (Seki *et al.*, 2004) and were started 50 ms after the end of image acquisitions (Fig. 1). The visual stimuli were presented at a visual angle of $5.5 \times 10.6^\circ$. We used an event-related design to minimize habituation and learning effects. The design consisted of four types of event conditions: auditory-auditory (AA), visual-visual (VV) and auditory-visual (AV) matching tasks, and a still face condition (STILL). For the AA, VV and AV matching tasks, both auditory and visual stimuli were presented in order to control for the sensory input. Through the session, the subjects were asked to fixate a small cross-hair at the center of the screen. We explicitly instructed the subjects not to close their eyes during the tasks except for blinking. During AA, an instruction cue was presented during the last 800 ms of the scan acquisition time (Fig. 1). Following a 1600 ms silent period, a single frame of two faces pronouncing a vowel (/a/, /e/, /i/, /o/ or /u/) was presented (Fig. 1). At the same time two consecutive voices pronouncing a vowel were played (e.g. /a/ followed by /e/). The timing of the first voice was synchronized with the faces. Immediately after the presentation, the cross-hair was changed to a minus sign, which indicated that the subject should press the right index finger button if the first and second pronounced vowels were the same, and otherwise

subjects were asked to press the middle finger button. The subjects were instructed to respond as quickly as possible, within 1600 ms. During VV, the same stimuli were presented but with different instructions, asking the subjects to compare the facial movements of the side-by-side stimuli. During the AV condition, subjects were asked to compare the first voice and the left face, or the first voice and the right face; hence, the synchronized stimuli were matched, excluding the effect of stimulus onset asynchrony (Bushara *et al.*, 2001). During the STILL condition, a face without any movement was presented for 1600 ms followed by the presentation of an arrow for 1600 ms; the arrow pointed right or left. The subject was asked to press the right index finger button if the leftward arrow was shown, and the middle finger button if the rightward arrow appeared. The inter-trial interval (ITI) was fixed at 4 s. Each condition was repeated 30 times, to give a total number of events of 120.

We wanted to maximize the efficiency with which we could detect differences between AA and STILL, VV and STILL, AV and STILL, AA and VV, AV and AA, and AV and VV. To do this, the distribution of the stimulus onset asynchrony (SOAs) of each condition were determined as follows (Friston *et al.* 1999b): the order of 120 events, 30 for each condition, was randomly permuted to generate a set of three vectors (1×120 matrix) indicating the presence (1) or absence (0) of particular event, and hence representing the distribution of the SOAs of each condition (SOA vectors). A design matrix incorporating four conditions (AA, VV, AV and STILL) was created by convolving a set of SOA vectors with a hemodynamic response function (h):

$$X = [aa, vv, av, s] \otimes b \quad X = [aa, vv, av, s] \otimes b$$

where aa represents AA, vv represents VV, av represents AV and s represents STILL.

The efficiency of the estimations of AA-STILL, VV-STILL, AV-STILL, AA-VV, AV-AA and AV-VV was evaluated using the inverse of the covariance of the contrast of the parameter estimates (Friston *et al.* 1999b):

$$\text{var}\{c^T \hat{\beta}\} = \sigma^2 c^T (X^T X)^{-1} c$$

$$\text{Efficiency} = \text{trace}\{c^T (X^T X)^{-1} c\}^{-1}$$

where $c = (1, 0, 0, -1)$ for AA-STILL, $(0, 1, 0, -1)$ for VV-STILL, $(0, 0, 1, -1)$ for AV-STILL, $(-1, 0, 1, 0)$ for AV-AA, $(1, -1, 0, 0)$ for AA and VV, and $(0, -1, 1, 0)$ for AV-VV. From the 100 000 randomly generated sets of SOA vectors, we selected the most efficient one, which showed a maximum of the sum of the squares of the efficiency vectors for six contrasts. Here we assume the error variance is constant (Mechelli *et al.*, 2003). The session was repeated three times.

Data Analysis

The first 4 vols of each fMRI session were discarded to allow for stabilization of the magnetization, and the remaining 120 vols per session (a total of 360 vols per subject) were used for analysis. The data were analyzed using statistical parametric mapping (SPM99, Wellcome Department of Cognitive Neurology, London, UK) implemented in Matlab (Mathworks, Sherborn, MA) (Friston *et al.*, 1994, 1995a,b). Head motion was corrected with the realignment program of SPM99 (Friston *et al.* 1995a). There was no trend of head motion correlated with the task. Following realignment, all images were coregistered to the high-resolution, 3-D, T1-weighted MRI, in reference to the anatomical T1-weighted MRI from locations identical to those in the fMRI images. The parameters for affine and nonlinear transformation into a template of T1-weighted images already fit to standard stereotaxic space (MNI template) (Evans *et al.*, 1994) were estimated with the high-resolution, 3-D, T1-weighted MR images by least square means (Friston *et al.*, 1995a,b). The parameters were applied to the coregistered fMRI data. The anatomically normalized fMRI data were filtered using a Gaussian kernel of 8 mm (full width at half maximum) in the x, y, and z axes.

Statistical Analysis

Statistical analysis in the present study was conducted at two levels. First, individual task-related activation was evaluated. Second, the summary data for each individual were incorporated into the second-level

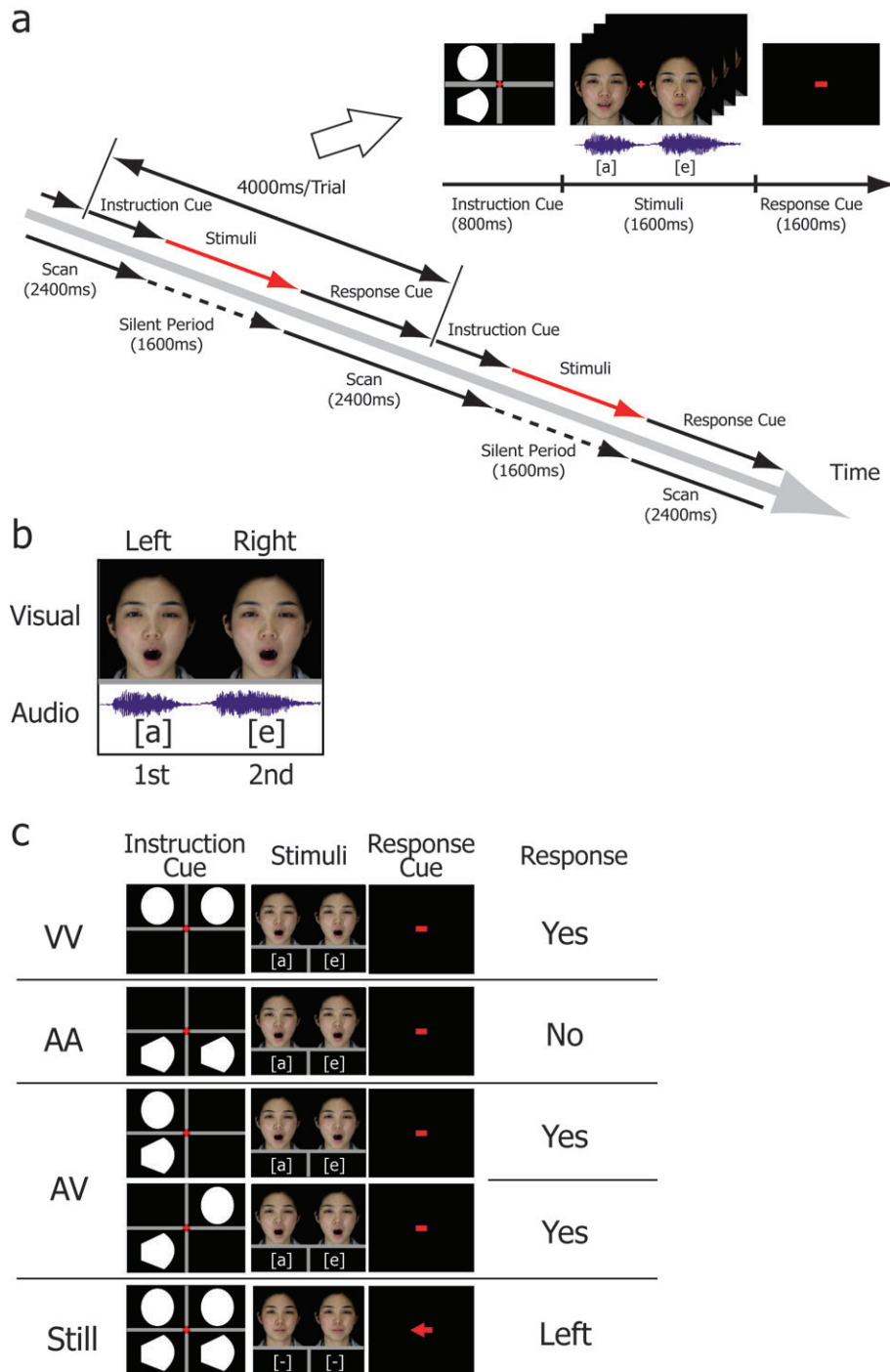


Figure 1. Sequence of the face-voice matching tasks and MR acquisition. (a) The time interval between two successive acquisitions of the same slice was 4000 ms. The cluster volume acquisition time was 2400 ms, leaving a 1600 ms silent period (red arrow) during which the face-voice stimuli were presented. (b) A single frame of two faces pronouncing a vowel was presented. At the same time two consecutive voices pronouncing a vowel such as /a/ followed by /e/ was provided. The first voice was synchronized in time with the faces. (c) Prior to the stimulus presentation, the instruction cue was provided for 800 ms. The instruction cue indicates which comparison should be performed. The subjects were requested to make the button press as soon as possible once the fixation marker (red hair cross) changed in shape to horizontal bars or arrows.

analysis using a random effect model (Friston *et al.*, 1999a) to make inferences at a population level.

Individual Analysis

The signal was scaled proportionally by setting the whole-brain mean value to 100 arbitrary units. The signal time course for each subject was modeled using a box-car function convolved with a hemodynamic response function, session effect, and high-pass filtering (112 s). The

explanatory variables were centered at 0. To test hypotheses about regionally-specific condition effects, the estimates for each of the model parameters were compared with the linear contrasts. First, we delineated the areas activated during the AA, VV and AV tasks compared with those activated during the STILL periods of the same session. AA-VV and VV-AA comparisons were conducted to depict the neural substrates of within-modal matching. Cross-modal areas were defined as those more prominently activated when there was matched information from

two different sensory modalities than when the matching was within either modality. Regions activated by both AV-AA and AV-VV were depicted by the intersection of the areas defined by each contrast. The statistical threshold of each contrast was set at $P < 0.05$, corrected for multiple comparisons at the cluster level for the entire brain. Within these cross-modal areas, we made a comparison between the AV sessions with concordant stimuli and those with discordant stimuli. The statistical threshold was set at $P < 0.05$, corrected for multiple comparisons within the limited search volume.

Group Analysis with the Random Effect Model

The weighted sum of the parameter estimates in the individual analysis constituted 'contrast' images, which were used for the group analysis (Friston *et al.*, 1999a). Contrast images obtained via individual analysis represented the normalized task-related increment of the MR signal of each subject. For each contrast, a one sample *t*-test was performed for each and every voxel within the brain to obtain population inferences. The resulting set of voxel values for each contrast constituted a statistical parametric map of the *t* statistic ($SPM\{t\}$). The $SPM\{t\}$ was transformed to normal distribution units ($SPM\{Z\}$). The threshold for $SPM\{Z\}$ was set at $Z > 3.09$ and $P < 0.05$, with a correction for multiple comparisons at the cluster level for the entire brain (Friston *et al.*, 1996). Similar to the individual analysis, cross-modal areas were depicted. Within these areas, we compared the AV sessions with concordant stimuli (AV congruent) and those with discordant stimuli (AV incongruent).

Results

Task Performance

The mean (\pm SD) percentage of correct responses was $84.4 \pm 8.2\%$ for the AV task, $90.0 \pm 6.0\%$ for the AA task, $86.3 \pm 8.5\%$ for the VV task and $98.5 \pm 2.5\%$ for the STILL condition. The performance on the STILL task was significantly better than on the other tasks, and the AA task performance was better than the AV [$P < 0.0001$, one-way analysis of variance (ANOVA) followed by Fisher's PLSD]. The mean (\pm SD) reaction time was 414 ± 109 ms for the AV task, 394 ± 110 ms for the AA task, 367 ± 111 ms for the VV task and 537 ± 108 ms for the STILL condition. The reaction times for the STILL task were significantly slower than those for the other tasks ($P < 0.0001$, one-way ANOVA followed by Fisher's PLSD).

Group Analysis with Random Effect Model (Figs 2-5, Tables 1-3)

As the left hander showed a activation pattern similar to those of the right-handed subjects, the group analysis was conducted with all subjects. Compared with rest, the auditory matching task activated the bilateral transverse (GTT) and superior temporal gyri (GTs), left inferior frontal gyrus (GFi), dorsal premotor cortex (PMd), inferior (LPi) and superior parietal lobule (LPs), fusiform gyrus (GF), the right cerebellum and lingual gyrus (GL), and the supplementary motor area (SMA). The visual matching task activated the bilateral occipito-temporal junction extending to the GL, GF, GTs, GFi, middle frontal gyri (GFm), LPi, LPs, left PMd, right cerebellum and anterior cingulate gyrus (ACG). The audio-visual matching task activated the bilateral GTs, PMd, LPi, LPs, occipito-temporal junction, cerebellum, left GFi, GFm, GTm, intraparietal sulcus (IPS), thalamus, right GL, GF, insula and ACG (Fig. 2).

Compared with the visual matching condition, the auditory matching task induced more prominent activation in the bilateral GTs and STS anterior to the Vpc line [an imaginary vertical line in the mid-sagittal plane passing through the anterior margin of the posterior commissure (Talairach and

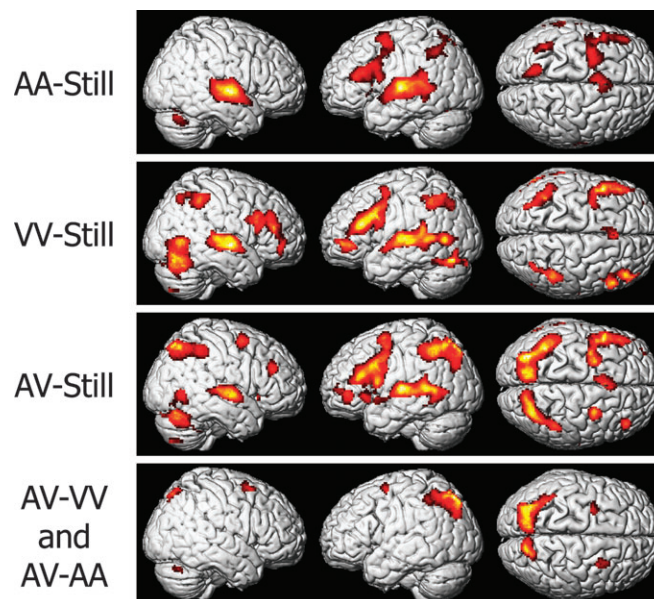


Figure 2. Group analysis of task-related activation by the contrasts of AA-STILL (top row), VV-STILL (middle row) and AV-STILL (bottom row). The activated areas ($P < 0.05$, corrected) are superimposed on surface-rendered high-resolution MRIs viewed from the right (left column), left (middle column) and top (right column).

Tournoux, 1988)], including the part of the frontal operculum and the left GTs extending to the STS posterior to the Vpc, and the left lingual gyrus (AA-VV). In these areas, the task-related increment during VV was similar to that during AV (Fig. 3). Compared with the auditory matching condition, the visual matching task activated the bilateral GFi and GFm, the right occipito-temporal junction extending to the GL and GF, the IPS, left cerebellum and ACG (VV-AA). In these areas, the task-related changes during the VV condition were similar to those during AV (Fig. 3).

More prominent activation during the AV task compared with both the AA and VV tasks was observed in the bilateral PMd and LPs, left IPS and right cerebellum. Within these areas, concordant AV stimuli activated the bilateral LPs and left IPS more prominently than discordant stimuli ($P < 0.05$, paired *t*-test with random effect model; Fig. 4 and Table 3). This congruency effect was not observed during AA or VV conditions. A typical individual dataset is presented to reveal the more prominent activation during concordant than discordant AV (Fig. 5).

The reverse contrasts (VV-AV masked with VV-STILL, and AA-AV masked with AA-STILL) were used to identify brain areas with significantly lower activity during cross-modal matching than during unimodal matching conditions (Table 4 and Fig. 6). During the AV condition compared with the AA condition, there was a decrease in signal in the bilateral primary and association auditory cortex, and the cerebellum. Compared with the VV condition, the activity in the right MT/V5 area was reduced during the AV condition (Fig. 6).

Discussion

Task Design

In the present study, we hypothesized that the cross-modal matching of voice and lip movements requires both divided attention and cross-modal binding.

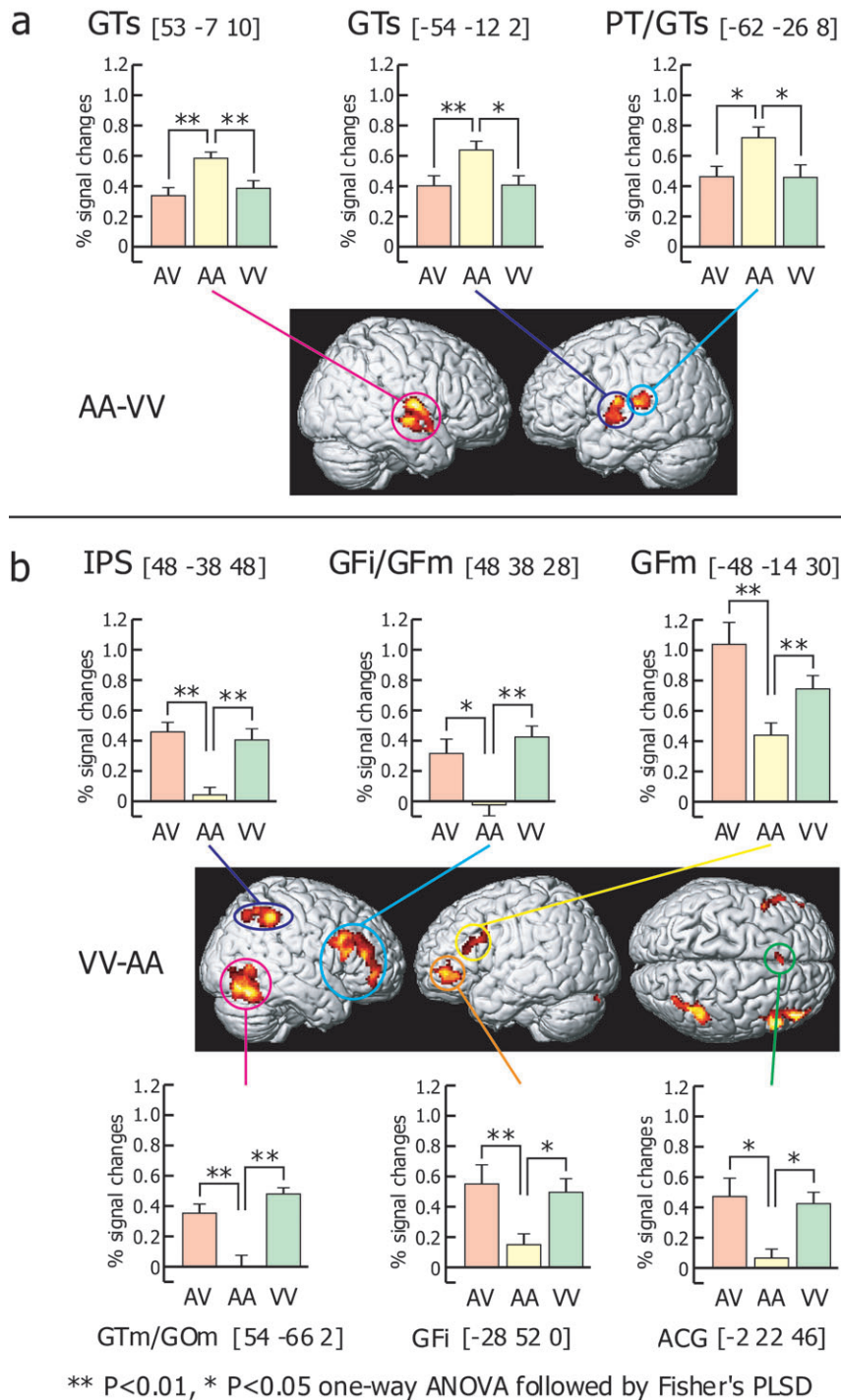


Figure 3. (a) Neural substrates of within-modal matching by comparing AA-VV. The more pronounced activation seen during AA than VV ($P < 0.05$ corrected) is superimposed on surface-rendered high-resolution MRIs viewed from the right (left of the second row) and left (right of the second row). Top and bottom rows show the task-related activation (% signal change) in each area during each condition. (b) Neural substrates of within-modal matching by VV-AA. The more pronounced activation seen during VV than AA ($P < 0.05$ corrected) is superimposed on surface-rendered high-resolution MRIs viewed from the right (left of the fourth row), left (middle of the fourth row) and top (right of the fourth row). The third and bottom rows show the task-related activation (% signal change) in each area during each condition. * $P < 0.05$, ** $P < 0.01$, one-way ANOVA followed by Fisher's PLSD.

Attention

To isolate the divided attention component, we explicitly targeted attentional modulation by preparing audio-visual stimuli in which subjects attend only to the appropriate modality. Stimuli were identical across the conditions, while the subject was required to shift their attention to each modality or across

modalities according to the cued instructions. The motor responses based on the unimodal and cross-modal comparisons were identical.

When recording brain activity during the tasks, different types of signals correspond to the activation of the attentional mechanism ('source' signals) and its interaction with the

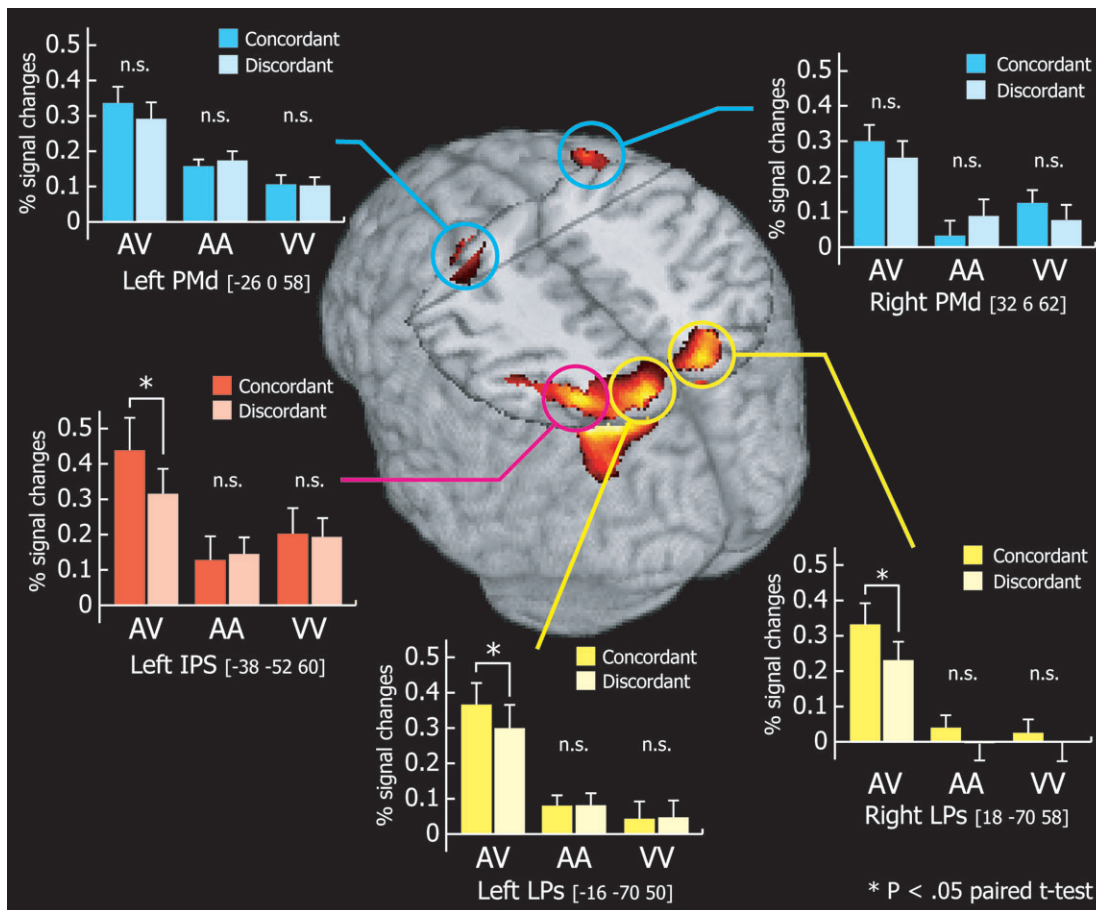


Figure 4. Statistical parametric maps of the group data. The neural substrates of cross-modal matching that were revealed by the intersection of AV-VV and AV-AA are superimposed on surface-rendered high-resolution MRIs. Bar graphs indicate the task-related activation (% signal change) by the concordant and discordant stimuli during the AV, AA and VV conditions in the left PMd, right PMd, left IPS, left LPs and right LPs using volume of interest with a 4 mm sphere. Error bars indicate the standard error mean. * indicates statistically significant difference ($P < 0.05$, paired t -test).

sensory systems ('site' signals) (Corbetta, 1998). In the present study, attention shifting is initiated by the instructions to the subject. A source signal would be associated with a modality-related shift of attention and would be recorded in areas that implement the attentional mechanism and/or in sensory areas responsible for stimulus analysis. During the VV task, for example, a source signal may prime visual processes for a more efficient response. Once a stimulus is presented, stimulus analysis may be enhanced by attention. This would produce modulation of visual processing ('site' signal), and this signal would mark the site of the interaction between the source attentional signals and visual processes (Corbetta, 1998). The same situation would occur in auditory processing during the AA task. Whereas source signals provide information regarding the organization of attention systems, site signals provide information on how sensory systems are affected by attention. Hence, VV-AA will show the differences in site signals representing the interaction of source attentional signals with visual processes. Similarly, AA-VV will reveal similar data for auditory processes. Here we assume that the same 'source' signals represent the attentional mechanisms in each condition, because they are directed to a single modality. On the other hand, during the AV task, compared with the unimodal tasks, the source signal would be recorded in areas that implement attention mechanisms, given that divided attention is more

demanding than unimodal attention, which may reflect the activation illustrated in Figure 4.

Cross-modal Binding

Additional neural processing during the AV condition compared with the unimodal VV or AA conditions includes cross-modal binding, in addition to cross-modal divided attention, arousal effects or other task-related effects, such as difficulty. In particular, the AV condition includes the switching of the direction of the spatial attention indicated by the visual cue, whereas AA and VV do not. Hence, the parieto-premotor cortex depicted by the AV-AA and AV-VV comparisons (Fig. 4) partly represents the neural substrates of switching spatial attention. This is concordant with the previous report indicating that switching the direction of the spatial attention is controlled by parieto-premotor cortical networks (Hopfinger *et al.*, 2000). Due to the methodological limitation, congruency cues in spatial location of the face and corresponding voice were missing. This might have biased one's perception and cross-modal processing mechanisms. To extract the component of cross-modal binding from these confounds, we utilized the cross-modal response enhancement by comparing a concordant stimulus with a discordant one. Based on an analogy with electrophysiological studies in nonhuman primates and other

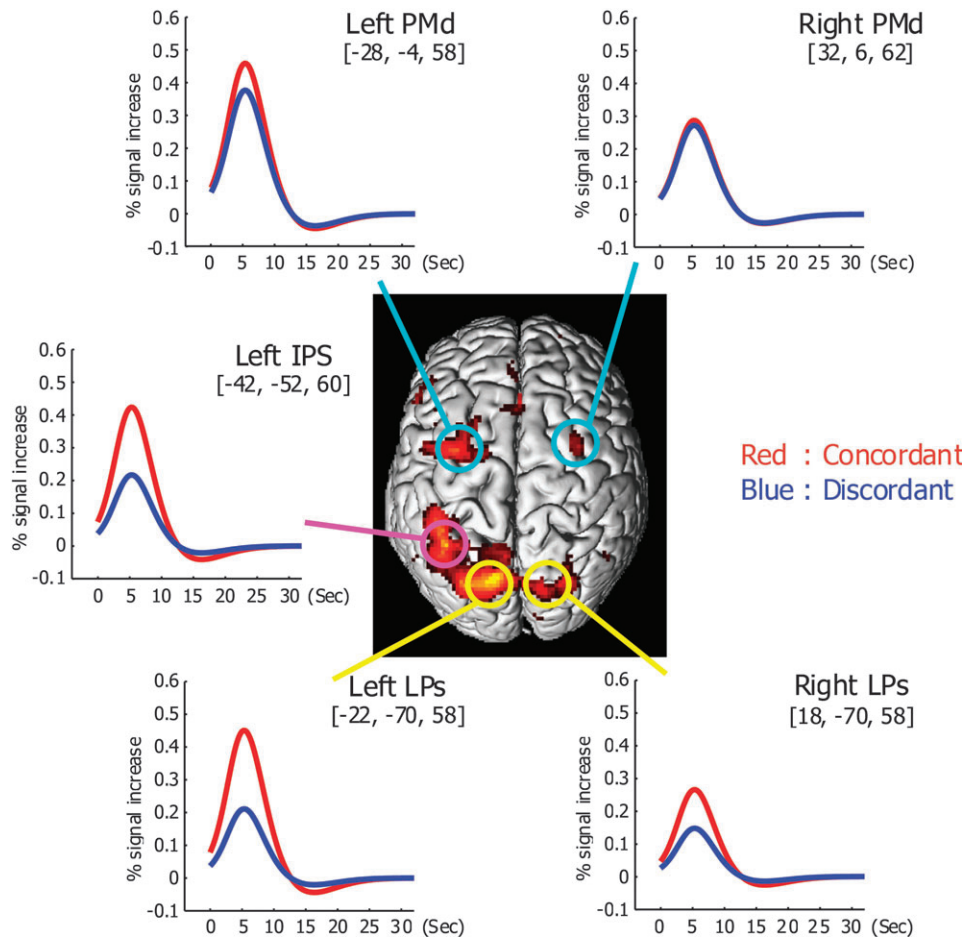


Figure 5. Statistical parametric maps from single subject. The neural substrates of cross-modal matching that were revealed by the intersection of AV-VV and AV-AA ($P < 0.05$ corrected) are superimposed on the surface-rendered high-resolution MRI of this subject (center). Event-related activation by AV congruent stimuli (red line) and AV incongruent stimuli (blue line) in the bilateral PMd and LPS and left LPI are shown.

mammals in the superior colliculus or cortex (Wallace *et al.*, 1992, 1996; King and Palmer, 1985), Calvert *et al.* (2000) postulated that response enhancement and depression are hallmarks of these intersensory interactions in humans.

Unimodal Matching

VV-AA

The VV-AA comparison revealed activation in the right occipito-temporal junction (hMT, the homologue of the simian MT) and LPI, which are part of the dorsal motion pathway, and bilateral prefrontal areas.

In monkeys, the cortical processing of visual motion is thought to involve anatomically inter-connected visual areas known as the dorsal motion pathway. These include lamina 4B in V1, the thick cytochrome oxidase stripes in V2, areas V3, MT (middle temporal visual area), MST (middle superior temporal visual area) and possibly the lateral and ventral intraparietal areas, LIP and VIP (Desimone and Ungerleider, 1986; Orban *et al.*, 1986; DeYoe and Van Essen, 1988; Boussaoud *et al.*, 1990). In humans, areas V1 and V2 are responsive to visual motion, but more selective responses can be obtained from extrastriate visual areas. For instance, hMT is strongly activated by visual motion stimuli and by visual motion discrimination tasks (Corbetta *et al.*, 1991; Zeki *et al.*, 1991; Dupont *et al.*, 1994;

Orban *et al.*, 1995; Tootell *et al.*, 1995a,b; Beauchamp *et al.*, 1997). Additionally, the same stimuli and tasks concurrently activate areas in dorsal occipital and posterior parietal cortex. Bilateral lesions of the lateral occipital cortex (including hMT) can selectively compromise visual motion perception, while leaving auditory and somatosensory motion perception intact (Zihl *et al.*, 1983, 1991; Rizzo *et al.*, 1995). Hence, along the dorsal visual stream, areas up to hMT may be specific to visual motion processing.

The activation in the occipito-temporal junction may include the posterior STS region that is adjacent to MT/V5. The posterior STS is known to be activated during the perception of human body movement (Bonda *et al.*, 1996; Howard *et al.*, 1996; Puce *et al.*, 1998).

A previous functional MRI study showed that mere observation of mouth actions activates bilateral ventral Brodmann's area (BA) 6 and 44, plus the right BA 45 in the inferior frontal gyrus (Buccino *et al.*, 2001). Other studies revealed activity in Broca's area during prepositional speech tasks, including reading words (Price *et al.*, 1994), word generation (McCarthy *et al.*, 1993), decoding syntactically complex sentences (Just *et al.*, 1996) and phoneme discrimination (Zatorre *et al.*, 1992, 1996). Auditory phonemic-discrimination activated secondary auditory cortices bilaterally, as well as Broca's area (Zatorre *et al.*, 1992, 1996). Zatorre *et al.* (1996) proposed that Broca's area is recruited

Table 1Task-related activation ($n = 17$)

Task	Cluster level		x (mm)	y (mm)	z (mm)	Z-value	Location	
	* P	Size					Side	Area (BA)
AA-STILL	<0.001	3174	-60	-16	6	6.44	Lt	GTs (42)
			-42	-19	11	6.06	Lt	GTT (41)
	<0.001	2713	51	-10	2	7.02	Rt	GTs (22)
			45	-21	10	5.80	Rt	GTT (41)
	<0.001	2206	-36	4	53	5.43	Lt	PMd (6)
			-32	27	15	5.27	Lt	GFi (45)
	<0.001	585	-36	19	10	4.46	Lt	GFi (44)
			-6	6	58	4.41	Lt	SMA (6)
	<0.001	573	-18	-67	-9	4.77	Lt	GF (19)
	<0.001	428	19	-65	-2	3.77	Rt	GL (18)
0.001	409	21	-69	-14	4.30	Rt	cerebellum	
0.001	397	-38	-49	42	4.85	Lt	LPi (40)	
0.002	333	-12	-62	54	5.19	Lt	LPs (7)	
VV-STILL	<0.001	3045	51	-65	1	5.13	Rt	GTm/GOm (37)
			45	-58	-16	5.04	Rt	GF (37)
	<0.001	2285	19	-73	-10	4.86	Rt	GL (18)
			-24	-73	-10	4.16	Lt	GL (18)
	<0.001	2236	-46	-58	-16	4.10	Lt	GF (37)
			37	-71	-17	4.99	Rt	cerebellum
	<0.001	1391	26	-66	-50	4.03	Rt	cerebellum
			-48	16	26	5.41	Lt	GFi 44
	<0.001	1216	-28	50	-1	4.61	Lt	GfM (10)
			-36	2	51	4.24	Lt	PMd (6)
<0.001	700	-62	-18	6	5.48	Lt	GTs (22)	
		-56	-62	6	4.86	Lt	GTm/GOm (37)	
<0.001	534	51	-10	-2	6.25	Rt	GTs (22)	
		45	48	-1	4.36	Rt	GfM (10)	
<0.001	534	-38	-49	41	4.51	Lt	LPi (40)	
		-28	-68	43	4.15	Lt	LPs (7)	
.008	289	45	-39	47	4.31	Rt	LPi (40)	
		37	-60	49	3.74	Rt	LPs (7)	
AV-STILL	<0.001	4749	-2	23	41	4.28	Lt	ACG (32)
			-60	-16	6	5.68	Lt	GTs (22)
	<0.001	2913	-48	16	26	5.13	Lt	GFi (44)
			-34	2	51	4.61	Lt	PMd (6)
	<0.001	2523	-56	-40	-9	3.18	Lt	GTm (21)
			-54	-61	3	3.76	Lt	GTm/GOm (37)
	<0.001	1385	-40	-49	42	5.11	Lt	LPi (40)
			-30	-66	40	6.52	Lt	IPS (40/7)
	<0.001	835	-12	-74	47	4.82	Lt	LPs (7)
			33	-58	-26	5.50	Rt	cerebellum
<0.001	835	-6	-79	-17	4.01	Lt	cerebellum	
		7	-81	-10	4.00	Rt	GL (18)	
<0.001	835	49	-65	-11	3.27	Rt	GF (19)	
		49	-63	1	4.30	Rt	GTm/GOm (37)	
<0.001	835	31	-62	50	5.99	Rt	LPs (7)	
		37	-53	44	4.70	Rt	LPi (40)	
<0.001	835	61	-16	2	4.31	Rt	GTs (22)	
<0.001	508	-4	17	43	5.76	Lt	ACG (32)	
0.001	387	-12	-12	9	3.28	Lt	thalamus	
0.007	268	31	6	53	4.50	Rt	PMd (6)	
0.013	232	-30	48	1	3.97	Lt	GfM (10)	
0.022	206	32	-25	-1	4.45	Rt	insula (13)	

ACG, anterior cingulate gyrus; BA, Brodmann's area; GF, fusiform gyrus; GFi, inferior frontal gyrus; GL, lingual gyrus; GOm, middle occipital gyrus; GTm, middle temporal gyrus; GTs, superior temporal gyrus; GTT, transverse temporal gyrus; IPS, intraparietal sulcus; LPi, inferior parietal lobule; LPs, superior parietal lobule; PMd, dorsal premotor cortex; SMA, supplementary motor cortex.

*With correction at the cluster level.

for fine-grained phonetic analysis by means of articulatory decoding. The phonetic analysis of speech depends not only on auditory information but also on access to information about the articulatory gestures associated with a given speech sound (Lieberman and Whalen, 2000). This access might be accomplished through a 'mirror' system, including Broca's area, which forms a link between the observer and the speaker (Rizzolatti and Arbib, 1998; Iacoboni *et al.*, 1999).

Table 2Within-modes activation ($n = 17$)

Task	Cluster level		x (mm)	y (mm)	z (mm)	Z-value	Location	
	* P	Size					Side	Area (BA)
AA-VV	<0.001	625	53	-7	10	4.84	Rt	GTs (22)
			57	-12	-3	4.36	Rt	STS (21/22)
	<0.001	278	-57	-7	11	4.42	Lt	GTs (22)
			211	-57	-30	16	4.06	Lt
VV-AA	0.001	184	-18	-58	-2	3.85	Lt	GL (19)
			111	-44	-25	-4	3.89	Lt
	<0.001	874	44	-65	-14	4.72	Rt	GF (19)
			24	-78	-3	3.99	Rt	GL (18)
<0.001	812	55	-60	3	3.73	Rt	GTm/GOm (37)	
		57	26	15	5.28	Rt	GFi (45)	
		44	50	-3	4.50	Rt	GfM (10)	
		449	42	-50	50	4.75	Rt	IPS (40/7)
<0.001	291	-49	48	-2	5.12	Lt	GFi (10)	
		136	-55	26	16	3.87	Lt	GFi (45)
		-44	13	34	3.25	Lt	GfM (9)	
		104	-14	-82	-16	3.69	Lt	cerebellum
0.041	85	-4	29	37	5.36	Lt	ACG (32)	

ACG, anterior cingulate gyrus; BA, Brodmann's area; GF, fusiform gyrus; GFi, inferior frontal gyrus; GfM, middle frontal gyrus; GL, lingual gyrus; GOm, middle occipital gyrus; GTm, middle temporal gyrus; GTs, superior temporal gyrus; IPS, intraparietal sulcus; LPi, inferior parietal lobule; LPs, superior parietal lobule; PMv, ventral premotor cortex; STS, superior temporal sulcus.

*With correction at the cluster level.

Table 3Cross-modal activation ($n = 17$)

Task	Cluster size	x (mm)	y (mm)	z (mm)	Z-value	Location		
						AV-AA	AV-VV	Side
AV-VV and AV-AA	1944	-38	-50	55	3.78	3.61	Lt	IPS (40/7)
		-24	-72	45	4.97	4.76	Lt	LPs (7)
		23	-62	50	4.70	3.10	Rt	LPs (7)
		29	-64	-26	3.98	3.81	Rt	cerebellum
		64	-26	4	58	4.24	5.30	Lt
84	31	16	54	4.12	4.03	Rt	PMd (6)	

Intersection of the areas defined by AV-AA and AV-VV at the threshold of $P < 0.05$ (cluster level). BA, Brodmann's area; IPS, intraparietal sulcus; LPs, superior parietal lobule; PMd, dorsal premotor cortex.

AA-VV

The AA-VV comparison activated the anterior portion of the GTs and STS bilaterally, and the posterior GTs corresponding to planum temporale (PT) and STS on the left side. Compared with visual motion processing, the neural substrates of audible speech processing are not as well understood. Activations along the STS are often reported in neuroimaging studies of human speech processing (Zatorre *et al.*, 1992; Dehaene *et al.*, 1997); however, their exact role is not clear. Belin *et al.* (2000) reported voice-selective regions located bilaterally along the upper bank of the superior temporal sulcus (STS) of humans ($x = 62$, $y = -14$, $z = 0$; left: $x = -58$, $y = -18$, $z = -4$, in Talairach's coordinates). They postulated that this area was the human homologue of TAA of the macaque (Seltzer and Pandya, 1978); this area forms part of a hierarchically organized system that is specialized for extracting auditory object features (Rauschecker, 1997; Kaas and Hackett, 1999). Belin *et al.* (2000) concluded that this area may be involved in the high-level analysis of complex acoustic information, such as the extraction of speaker-related cues, and transmission of this information to other areas for multimodal integration (Mesulam, 1998).

Table 4
Cross-modal deactivation ($n = 17$)

Task	Cluster level		x (mm)	y (mm)	z (mm)	Z-value	Location	
	* P	Size					Side	Area (BA)
AA-AV	<0.001	1752	60	4	0	4.78	Rt	GTs (22)
	<0.001	1015	-58	-4	2	4.97	Lt	GTs (22)
	0.019	106	-10	-64	-10	4.66	Lt	cerebellum
WV-AV	0.012	144	52	-70	2	4.16	Rt	GTm/GOm (37)

BA, Brodmann's area; GOm, middle occipital gyrus; GTm, middle temporal gyrus; GTs, superior temporal gyrus.

*With correction at the cluster level.

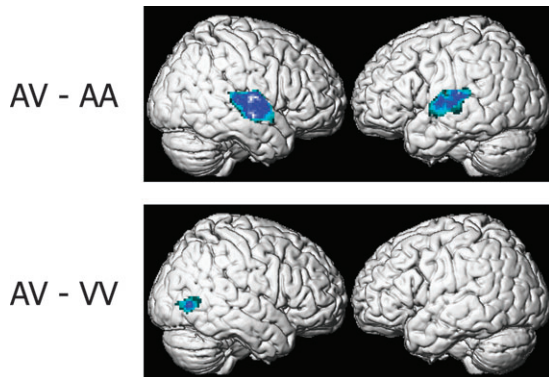


Figure 6. Areas with lower activity during cross-modal matching (AV) compared with unimodal matching (AA, top row; WV, bottom row). The areas are superimposed on the surface rendered high resolution MRIs viewed from right (left column) and left (right column). The statistical threshold is $P < 0.05$, corrected.

The left-lateralized activity in the PT and adjacent STS might be related to the notable anatomical and functional asymmetry seen in auditory processing. Anatomically, the PT is significantly larger in the left hemisphere, with increased cell size and density of neurons. Furthermore, there are more functionally distinct columnar systems per surface unit in the left than the right PT (Galuske *et al.*, 2000). Functionally, the left auditory cortical areas that are optimal for speech discrimination, which is highly dependent on rapidly changing broadband sounds, have a higher degree of temporal sensitivity. By contrast, the right cortical counterpart has greater spectral sensitivity, and thus is optimal for processing the tonal patterns of music, in which small and precise changes in frequency are important (Zatorre *et al.*, 2002).

Cross-modal Matching

The present study showed that cross-modal matching tasks activated the LPs and dorsal premotor cortex bilaterally, the left IPS and the right cerebellum compared with the unimodal matching tasks. Furthermore, the LPs and IPS revealed a cross-modal response enhancement: congruent stimuli (when the sound of the voice and movement of the lips represented the same vowel) activated these areas more prominently than the incongruent stimuli (congruency effect). The congruency effect was not observed during the AA or VV conditions (Fig. 4), and hence this response is AV-specific. As audible speech was combined with the visible articulation movements, we anticipated the interaction of visual motion processes and auditory speech processing.

The areas in and around the posterior IPS (IPP) are known to be polymodal areas. In macaque monkeys, the ventral intraparietal area (VIP), located in the fundus of the IPS, is known to contain cells with distinct polysensory receptive fields. The VIP is known to receive direct projections from motion-related visual areas MT, MST and the surrounding polymodal cortex, as well as from auditory-related cortices (Maunsell and Van Essen, 1983; Lewis and Van Essen, 2000); hence this area is part of the neural substrates for visual motion processing. Individual neurons in the lateral intraparietal area, LIP, respond selectively to the locations of both visual and auditory targets (Stricanne *et al.*, 1996). The neurons in VIP and LIP can represent visuospatial information in a frame of reference that is non-retinotopic (e.g. head- or world-centered) (Duhamel *et al.*, 1998; Snyder *et al.*, 1998). Furthermore, using fMRI, Bremner *et al.* (2001) revealed that the depth of the human IPS was equivalent to that of the monkey VIP using polysensory (visual, auditory and somatosensory) stimuli conveying motion information. These results strongly indicate that the cortical areas in and around the IPP are polysensory areas (Lewis *et al.*, 2000).

The present findings also suggest that the neural substrates of cross-modal binding are closely related to those for divided attention. This also has been implicated in tasks involving cross-modal spatial attention (Bushara *et al.*, 1999; Eimer, 1999). Using audio-visual cross-modal motion discrimination, Lewis *et al.* (2000) found activation in the IPS and dorsal premotor cortex. Lewis *et al.* (2000) suggested that the control of attention and the motion computations themselves may be intimately intertwined by common mechanisms within the fronto-parietal networks. Also, it has been proposed that the LPs are involved in the orienting of attention (Corbetta *et al.*, 1993).

The parietal-premotor cortical network is related to top-down attentional modulation (Hopfinger *et al.*, 2000). The region of the LPI close to the IPS is important for attention shifting whereas the LPs are active both during shifts of attention in the visual periphery and also when attention is focused on target stimuli and no attentional shifts are required (Hopfinger *et al.*, 2000; Corbetta, 1998). They are part of a network that is important for the control of attention (Driver and Spence, 1998; Mesulam, 1998). Numerous studies have reported that the parietal-premotor network is activated when attention is directed to vision (Posner *et al.*, 1987; Corbetta *et al.*, 1993; LaBar *et al.*, 1999), audition (Pugh *et al.*, 1996; Binder *et al.*, 1997), and cross-modal stimuli (O'Leary *et al.*, 1997; Bushara *et al.*, 1999). The parietal-premotor network may act to direct attention to targets within the same or different sensory modalities (Lewis *et al.*, 2000).

The present study revealed that the neural substrates of divided attention and cross-modal integration overlap in the IPP and LPs. This is consistent with the idea that cross-modal areas allow multidimensional integration through two interactive processes (Mesulam, 1998): the formation of a directory pointing to the distributed sources of the related information; and the establishment, by local neuronal groups, of convergent cross-modal associations related to a target event. Divided attention may enhance the process of the former, while cross-modal response enhancement (congruency effect) is related to the latter. Cross-modal IPP and the adjacent LPs thus enable the binding of modality-specific information into multimodal representations that have both distributed and convergent components.

The IPP was also active during tactile–visual cross-modal shape matching (Saito *et al.*, 2003), audio-visual motion speed discrimination tasks (Lewis *et al.*, 2000) and even olfactory-visual integration (Gottfried and Dolan, 2003), although the locations slightly differed from the present study. Considering this lack of specificity of IPP activation, it is unlikely that the IPP contains the amodal representations of speech itself. Instead, the IPP and adjacent LPs may represent a node through which the senses can access each other directly from their sensory-specific systems.

Finally, during AV condition, activities in the unimodal areas were reduced whereas multimodal areas were enhanced. This is consistent with the previous study (Bushara *et al.*, 2003), suggesting that a reciprocal and ‘competitive’ interaction between multimodal and unimodal areas underlies the integration of simultaneous signals from different sensory modalities.

Notes

This study was supported by a Grant-in Aid for Scientific Research B#14380380 (N.S.) from the Japan Society for the Promotion of Science, and Special Coordination Funds for Promoting Science and Technology from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

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

References

- Beauchamp MS, Cox RW, DeYoe EA (1997) Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *J Neurophysiol* 78:516–520.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309–312.
- Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17:353–362.
- Bonda E, Petrides M, Ostry D, Evans A (1996) Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J Neurosci* 16:3737–3744.
- Boussaoud D, Ungerleider LG, Desimone R (1990) Pathways for motion analysis: cortical connections of the middle superior temporal and fundus of the superior temporal visual areas in the macaque. *J Comp Neurol* 296:462–495.
- Bremmer F, Schlack A, Shah NJ, Zafiris O, Kubischik M, Hoffmann K, Zilles K, Fink GR (2001) Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron* 29:287–296.
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fagossi L, Gallese V, Seitz R, Zilles K, Rizzolatti G, Freund H-J (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur J Neurosci* 13:400–404.
- Bushara KO, Weeks RA, Ishii K, Catalan M-J, Tian B, Rauschecker JP, Hallett M (1999) Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat Neurosci* 2:759–766.
- Bushara KO, Grafman J, Hallett M (2001) Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci* 21:300–304.
- Bushara KO, Hanakawa T, Immisch I, Toma K, Kansaku K, Hallett M (2003) Neural correlates of cross-modal binding. *Nat Neurosci* 6:190–195.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. *Science* 276:593–596.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of cross-modal binding in the human heteromodal cortex. *Curr Biol* 10:649–657.
- Corbetta M (1998) Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? *Proc Natl Acad Sci USA* 95:831–838.
- Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1991) Selective and divided attention during visual discrimination of shape, color, and speed: functional anatomy by positron emission tomography. *J Neurosci* 11:2383–2402.
- Corbetta M, Miezin FM, Shulman GL, Petersen SE (1993) A PET study of visuospatial attention. *J Neurosci* 13:1202–1226.
- Dehaene S, Dupoux E, Mehler J, Cohen L, Paulesu E, Perani D, van de Moortele PF, Lehericy S, Le Bihan D (1997) Anatomical variability in the cortical representation of first and second language. *Neuroreport* 8:3809–3815.
- Desimone R, Ungerleider LG (1986) Multiple visual areas in the caudal superior temporal sulcus of the macaque. *J Comp Neurol* 248:164–189.
- DeYoe EA, Van Essen DC (1988) Concurrent processing streams in monkey visual cortex. *Trends Neurosci* 11:219–226.
- Driver J, Spence C (1998) Attention and the cross-modal construction of space. *Trends Cogn Sci* 2:254–262.
- Duhamel J-R, Colby CL, Goldberg ME (1998) Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *J Physiol* 79:126–136.
- Dupont P, Orban GA, De Bruyn B, Verbruggen A, Mortelmans L (1994) Many areas in the human brain respond to visual motion. *J Neurophysiol* 72:1420–1424.
- Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM (1999) Improved auditory cortex imaging using clustered volume acquisition. *Hum Brain Mapp* 7:89–97.
- Eimer M (1999) Can attention be directed to opposite locations in different modalities? An ERP study. *Clin Neurophysiol* 110:1252–1259.
- Evans AC, Kamber M, Collins DL, MacDonald D (1994) An MRI-based probabilistic atlas of neuroanatomy. In: *Magnetic resonance scanning and epilepsy* (Shorvon SD, ed.), pp. 263–274. New York: Plenum Press.
- Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC (1994) Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp* 1:210–220.
- Friston KJ, Ashburner J, Frith CD, Heather JD, Frackowiak RSJ (1995a) Spatial registration and normalization of images. *Hum Brain Mapp* 2:165–189.
- Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ (1995b) Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* 2:189–210.
- Friston KJ, Holmes A, Poline J-B, Price CJ, Frith CD (1996) Detecting activations in PET and fMRI: levels of inference and power. *Neuroimage* 4:223–235.
- Friston KJ, Holmes AP, Worsley KJ (1999a) How many subjects constitute a study? *Neuroimage* 10:1–5.
- Friston KJ, Zarahn E, Josephs O, Henson RN, Dale AM (1999b) Stochastic designs in event-related fMRI. *Neuroimage* 10:607–619.
- Galuske RAW, Schlote W, Bratzke H, Singer W (2000) Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science* 289:1946–1949.
- Gottfried JA, Dolan RJ (2003) The nose smells what the eye sees: cross-modal visual facilitation of human olfactory perception. *Neuron* 39:375–386.
- Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. *Nat Neurosci* 3:284–291.
- Howard RJ, Brammer M, Wright I, Woodruff PW, Bullmore ET, Zeki S (1996) A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. *Curr Biol* 6:1015–1019.
- Iacoboni M, Woods RP, Brass M, Becckering H, Mazziotta JC, Rizzolatti G (1999) Cortical mechanisms of human imitation. *Science* 286:2526–2528.
- Jones JA, Callan DE (2003) Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport* 14:1129–1133.

- Just MA, Carpenter PA, Keller TA, Eddy WF, Thulborn KR (1996) Brain activation modulated by sentence comprehension. *Science* 274:114-116.
- Kaas JH, Hackett TA (1999) 'What' and 'where' processing in auditory cortex. *Nat Neurosci* 2:1045-1047.
- King AJ, Palmer AR (1985) Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Exp Brain Res* 60:492-500.
- LaBar KS, Gitelman DR, Parrish TB, Mesulam M-M (1999) Neuroanatomic overlap of working memory and spatial attention networks: a functional MRI comparison within subjects. *Neuroimage* 10:695-704.
- Lewis JW, Van Essen DC (2000) Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J Comp Neurol* 428:112-137.
- Lewis JW, Beauchamp MS, Deyoe EA (2000) A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb Cortex* 10:873-888.
- Liberman AM, Whalen DH (2000) On the relation of speech to language. *Trends Cogn Sci* 4:187-196.
- Maunsell JH, Van Essen DC (1983) The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J Neurosci* 3:2563-2586.
- McCarthy G, Blamire AM, Rothman DL, Gruetter R, Shulman RG (1993) Echo-planar magnetic resonance imaging studies of frontal cortex activation during word generation in humans. *Proc Natl Acad Sci USA* 90:4952-4956.
- Mechelli A, Price CJ, Henson RNA, Friston KJ (2003) Estimating efficiency a priori: a comparison of blocked and randomized designs. *Neuroimage* 18:798-805.
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7:3215-3229.
- Mesulam MM (1998) From sensation to cognition. *Brain* 121:1013-1052.
- O'Leary DS, Andreasen NC, Hurtig RR, Torres IJ, Flashman LA, Kesler ML, Arndt SV, Cizadlo TJ, Ponto LLB, Watkins GL, Hichwa RD (1997) Auditory and visual attention assessed with PET. *Hum Brain Mapp* 5:422-436.
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97-113.
- Olson IR, Gatenby JC, Gore JC (2002) A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cogn Brain Res* 14:129-138.
- Orban GA, Kennedy H, Bullier J (1986) Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. *J Neurophysiol* 56:462-480.
- Orban GA, Dupont P, De Bruyn B, Vogels R, Vandenberghe R, Mortelmans L (1995) A motion area in human visual cortex. *Proc Natl Acad Sci USA* 92:993-997.
- Posner MI, Walker JA, Friedrich FA, Rafal RD (1987) How do the parietal lobes direct covert attention? *Neuropsychologia* 25:135-145.
- Price CJ, Wise RJ, Watson JD, Patterson K, Howard D, Frackowiak RS (1994) Brain activity during reading. The effects of exposure duration and task. *Brain* 117:1255-1269.
- Puce A, Allison T, Bentin S, Gore JC, McCarthy G (1998) Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci* 18:2188-2199.
- Pugh KR, Shaywitz BA, Shaywitz SE, Fulbright RK, Byrd D, Skudlarski P, Shankweiler DP, Katz L, Constable RT, Fletcher J, Lacadie C, Marchione K, Gore JC (1996) Auditory selective attention: an fMRI investigation. *Neuroimage* 4:159-173.
- Rauschecker JP (1997) Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngol Suppl* 532:34-38.
- Rizzo M, Mawrot M, Zihl J (1995) Motion and shape perception in cerebral akinetopsia. *Brain* 118:1105-1127.
- Rizzolatti G, Arbib A (1998) Language within our grasp. *Trends Neurosci* 21:188-194.
- Saito DN, Okada T, Morita Y, Yonekura Y, Sadato N (2003) Tactile-visual cross-modal shape matching: a functional MRI study. *Cogn Brain Res* 17:14-25.
- Seki A, Okada T, Koeda T, Sadato N (2004) Phonemic manipulation in Japanese: an fMRI study. *Cogn Brain Res* 20:261-272.
- Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res* 149:1-24.
- Snyder LH, Grieve KL, Brotchie P, Andersen RA (1998) Separate body- and world-referenced representations of visual space in parietal cortex. *Nature* 394:887-891.
- Stricane B, Andersen RA, Mazzoni P (1996) Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *J Neurophysiol* 76:2071-2076.
- Sumbly WH, Polack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212-215.
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.
- Tootell RBH, Reppas JB, Dale AM, Look RB, Sereno MI, Malach R, Brady TJ, Rosen BR (1995a) Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature* 375:139-141.
- Tootell RBH, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW (1995b) Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J Neurosci* 15:3215-3230.
- Wallace MT, Meredith MA, Stein BE (1992) Integration of multiple sensory modalities in cat cortex. *Exp Brain Res* 91:484-488.
- Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol* 76:1246-1266.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846-849.
- Zatorre RJ, Meyer E, Gjedde A, Evans AC (1996) PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cereb Cortex* 6:21-30.
- Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. *Trends Cogn Sci* 6:37-46.
- Zeki S, Watson JDG, Lueck CJ, Friston KJ, Kennard C, Frackowiak RSJ (1991) A direct demonstration of functional specialization in human visual cortex. *J Neurosci* 11:641-649.
- Zihl J, Von Cramon D, Mai N (1983) Selective disturbance of movement vision after bilateral brain damage. *Brain* 106:313-340.
- Zihl J, Von Cramon D, Mai N, Schmid CH (1991) Disturbance of movement vision after bilateral posterior brain damage. *Brain* 114:2235-2252.