Cerebral Cortex May 2013;23:1097–1107
doi:10.1093/cercor/bhs076
Advance Access publication April 19, 2012

Multisensory Interactions between Auditory and Haptic Object Recognition

Tanja Kassuba[1,2,3], Mareike M. Menz[2], Brigitte Röder[4] and Hartwig R. Siebner[1,2,3]

[1]Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital Hvidovre, 2650 Hvidovre, Denmark, [2]NeuroimageNord/Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany, [3]NeuroimageNord/Department of Neurology, Christian-Albrechts-University, 24105 Kiel, Germany and [4]Biological Psychology and Neuropsychology, University of Hamburg, 20146 Hamburg, Germany

Address correspondence to Tanja Kassuba, Danish Research Centre for Magnetic Resonance, Section 340B, Copenhagen University Hospital Hvidovre, Kettegaard Allé 30, 2650 Hvidovre, Denmark. Email: tanjak@drcmr.dk.

Object manipulation produces characteristic sounds and causes specific haptic sensations that facilitate the recognition of the manipulated object. To identify the neural correlates of audio-haptic binding of object features, healthy volunteers underwent functional magnetic resonance imaging while they matched a target object to a sample object within and across audition and touch. By introducing a delay between the presentation of sample and target stimuli, it was possible to dissociate haptic-to-auditory and auditory-to-haptic matching. We hypothesized that only semantically coherent auditory and haptic object features activate cortical regions that host unified conceptual object representations. The left fusiform gyrus (FG) and posterior superior temporal sulcus (pSTS) showed increased activation during crossmodal matching of semantically congruent but not incongruent object stimuli. In the FG, this effect was found for haptic-to-auditory and auditory-to-haptic matching, whereas the pSTS only displayed a crossmodal matching effect for congruent auditory targets. Auditory and somatosensory association cortices showed increased activity during crossmodal object matching which was, however, independent of semantic congruency. Together, the results show multisensory interactions at different hierarchical stages of auditory and haptic object processing. Object-specific crossmodal interactions culminate in the left FG, which may provide a higher order convergence zone for conceptual object knowledge.

Keywords: auditory perception, functional magnetic resonance imaging, multisensory interactions, object recognition, touch perception

Introduction

The visual appearance, haptic properties (including shape and surface), and characteristic sounds associated with the manipulation of an object provide both redundant and complementary features that facilitate object recognition. Several functional magnetic resonance imaging (fMRI) studies have investigated the neural correlates that are implicated in the integration of object-specific sensory input across vision and touch and across vision and audition (for reviews, see Amedi et al. 2005; Beauchamp 2005). For instance in the ventral visual pathway, the lateral occipital cortex (LO) processes visual and haptic object-related information when individuals grasp or see man-made tools, animal models, or toys (e.g., Amedi et al. 2001, 2002). On the other hand, the posterior superior temporal sulcus (pSTS) and adjacent middle temporal gyrus have been implicated in audio-visual integration of object-specific input across a wide range of living (e.g., voices/faces, animals) or nonliving objects (e.g., tools, weapons, musical instruments; Beauchamp, Argall, et al. 2004; Beauchamp, Lee, et al. 2004; Sestieri et al. 2006).

While previous neuroimaging work on multisensory object recognition has mainly focused on visuo-haptic and audio-visual integration, little is known about the convergence of object processing across audition and touch. Studies using low-level auditory and tactile stimuli (e.g., vibrotactile stimuli and related sounds) have implicated audio-tactile integration mechanisms in the posterior superior temporal gyrus (pSTG) and adjacent pSTS (Foxe et al. 2002; Schurmann et al. 2006; Beauchamp et al. 2008) as well as the anterior insula (Renier et al. 2009). Consistent with these results, data from a recent high-density electroencephalography (EEG) study using a haptic-to-auditory priming paradigm with meaningful object stimuli suggests integration-related neuronal activity in the left lateral temporal cortex (Schneider et al. 2011). In line with these data, we have recently shown in a fMRI study that the left fusiform gyrus (FG) and pSTS preferentially process manipulable object stimuli relative to nonobject control stimuli within both the auditory and the tactile modalities (Kassuba et al. 2011). These data showed that both regions were also activated during actual multisensory audio-haptic object processing (i.e., when the objects were presented simultaneously in both modalities). However, the contribution from either modality within these regions to higher order aspects of object recognition remains unclear.

Here, we employed event-related fMRI to study audio-haptic interactions in higher order object recognition. We opted for a delayed-match-to-sample paradigm in which participants had to match object-related sensory input within and across audition and touch. We manipulated the semantic congruency of sample and target stimuli to identify brain regions that are selectively responsive to semantically congruent crossmodal input. We reasoned that these cortical regions host unified object representations and, thus, subserve the binding of auditory and haptic input that refers to the same object concept (cf. Schneider et al. 2011). We chose familiar real manipulable objects and typical sounds created by a manipulation on them as sample and target stimuli because audio-haptic interactions are especially relevant in the context of object manipulation. Based on our previous results, we defined the left FG and pSTS as candidate regions to host unified audio-haptic object representations. We expected that crossmodal matching effects in these association regions would critically depend on the semantic congruency between sample and target objects.

Materials and Methods

Participants

Twenty-one healthy young volunteers (10 females) participated in the fMRI study. The age of the participants ranged from 21 to 33 years

(M ± standard deviation [SD]: 26.52 ± 2.93). Handedness was tested with the short form of the Edinburgh Handedness Inventory (Oldfield 1971); all participants were consistently right handed with a Laterality Index $\geq$ 0.87 (scaling adapted from Annett 1970). All participants had normal or corrected-to-normal vision, normal hearing ability, and normal tactile acuity, and none had a history of psychiatric or neurological disorders (self-report). Each individual gave written informed consent. The study was approved by the local ethics committee (Ärztekammer Hamburg) and in accordance with the Declaration of Helsinki.

Stimuli

Stimuli consisted of 24 different objects. These objects were man-made objects, including tools, toys, and musical instruments. All objects were fully graspable with one hand. Each object was associated with one characteristic sound derived from a manipulative action on it (e.g., a rubber duck was connected to its typical squeak). All sounds were cut to clips of 2 s of duration and equated for root-mean-square power using MATLAB (The MathWorks, Natick, MA).

Experimental Design and Procedure

We adopted an event-related fMRI paradigm. Examples of the different trials are depicted in Figure 1. Each trial consisted of a sample object stimulus (S1) and a target object stimulus (S2) presented successively. Upon presentation of S2, participants had to indicate whether the object pairs were either semantically congruent (50%) or incongruent (50%), that is, representing conceptually the same object or different objects. The object stimuli could be presented either haptically (actively palpating the object) or acoustically (hearing a typical sound of an object), and S1 and S2 were either of the same modality (unimodal) or of the different modalities (crossmodal). This resulted in a 2 × 2 × 2 factorial design with 2 possible modalities of S2 (auditory and haptic), 2 levels of semantic congruency (congruent and incongruent), and 2 levels of sensory matching (unimodal and crossmodal).

S1 and S2 were both presented for 2 s and preceded by a low-pitched (550 Hz) or high-pitched (800 Hz) tone or a double tone, respectively, to announce the sensory modality of the forthcoming stimulus. A single tone lasted for 300 ms and indicated S1, a double tone consisted of 2 short tones of 120 ms each with an intertone interval of 60 ms and indicated S2. After the instruction tone, there was a short pause of 100 ms before the stimulus was presented. The interstimulus and intertrial intervals (i.e., from the offset of a stimulus to the onset of the next indicating tone) varied between 2 and 6 s, jittered in steps of 1 s. Trials were presented in pseudorandomized order so that the same objects would not occur in successive trials and that the S1–S2 sensory modality combination would be repeated maximally once in successive trials. Every object appeared once as S1 and once as S2 in each condition, the combination of S1 and S2 objects in incongruent trials was randomized.

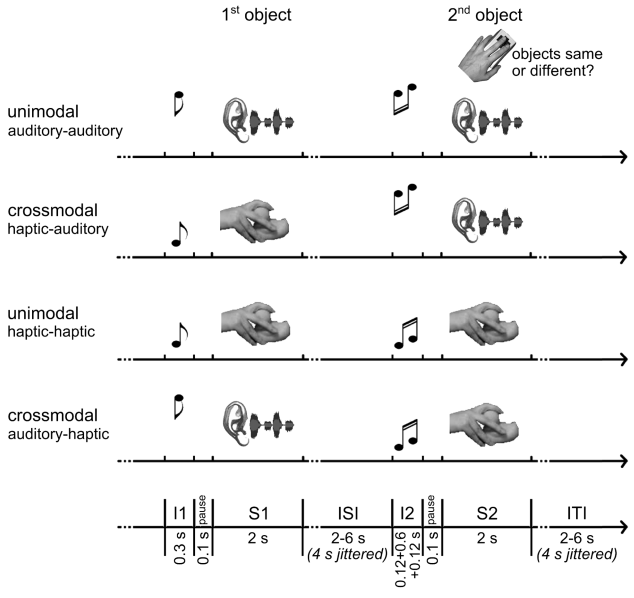

Figure 1. Illustration of the experimental design. Each trial consisted of a sample (S1) and a target object stimulus (S2), and participants had to decide by button press whether the object pairs were semantically congruent (50%) or incongruent (50%). Sample and target were either haptic or auditory stimuli and could both be presented either within the same modality (unimodal) or across modalities (crossmodal). A short high- or low-pitched tone (I1) or double tone (I2) preceding the stimuli informed the participants about the sensory modality S1 or S2, respectively, would be presented in. ISI = interstimulus interval, ITI = intertrial interval.

It is important to note that the participants did not know whether the S2 would be presented as an auditory or a haptic stimulus until 400 ms before onset of S2. At this point in time, an instruction tone informed the participants about the modality of S2. This means that all trials with an auditory S1 and all trials with a haptic S1 were identical, respectively, until shortly before the onset of S2. Thus, during the delay period, participants had to prepare for potential processing of both an auditory and a haptic object stimulus. Therefore, it is very likely that S1 was indeed encoded up to the semantic level.

One day before the fMRI experiment, participants were trained to recognize the 24 object stimuli by hearing the related sound and by haptic exploration (without ever seeing the objects). The training was repeated until the object stimuli were identified with an accuracy of 100% (0–2 repetitions per participant). Then, they were familiarized with the experimental task and trained to haptically explore the objects with an appropriate speed. A short recall of the training immediately before scanning showed that the objects were still identified with an accuracy of 100% within each modality. Throughout the whole experiment, participants were not able to hear the sound that their haptic exploration produced. In addition, participants were blindfolded during the training as well as during the fMRI experiment.

The event-related fMRI experiment consisted of 192 trials (24 trails per condition), separated into 2 runs of scanning of about 20 min (96 pseudorandomized trials per run). Haptic stimuli were placed on a board which was fixed onto the participants' waist by a vacuum cushion. Participants were asked to rest their right hand on the board and were instructed to wait for a presented tone. In case of a low-pitched tone, they were asked to make a slight movement to the left, palpate the haptic stimulus, and then to repose their hand. Participants were trained to use no more than 2 s for hand movements and exploration. The participants' right upper arm was completely fixed in a slightly elevated position and supported by cushions to minimize arm movements during haptic exploration. In case of a high-pitched tone, they were asked to wait for the auditory presentation of the following object. Upon the presentation of S2, participants were instructed to indicate by button press as fast and as accurately as possible whether both objects were the same or different. Responses were given by the middle and the index finger of the left hand, using an MR-compatible button device. The finger-response assignment was counterbalanced across participants, and the responses were recorded with Presentation software (Neurobehavioral Systems, Albany, CA).

Auditory stimuli were presented using Presentation running on a Windows XP professional SP3 PC. During MR image acquisition, they were presented at approximately 75 dB using an MR-compatible headphone system with gradient noise suppression (MR confon GmbH, Magdeburg, Germany). Sound volume was adjusted individually before the beginning of the first experimental run. Participants reported no problems with hearing the sound stimuli during MR image acquisition.

Magnetic Resonance Imaging

MRI measurements were carried out on a 3-T MRI scanner with a 12-channel head coil (TRIO, Siemens, Erlangen, Germany). To measure task-related changes in the blood oxygen level–dependent (BOLD) signal, we acquired 38 transversal slices (216 mm field of view [FOV], 72 × 72 matrix, 3 mm thickness, no spacing) covering the whole brain using a fast gradient echo T_2^*-weighted echo planar imaging sequence (repetition time [TR] 2480 ms, echo time [TE] 30 ms, 80° flip angle). Two functional runs with 505 ± 1 brain volumes each were acquired with each volume consisting of 38 slices. The exact number of volumes per run depended on the individual randomization of the jittered time

intervals between stimuli. During the first 5 and last 4 volumes of each run, participants fixated a cross without performing a task. In addition to fMRI, a high-resolution T_1-weighted anatomical volume was recorded for each participant, using a magnetization-prepared rapid acquisition gradient echo sequence (256 mm FOV, 256 × 192 matrix, 240 transversal slices, 1 mm thickness, 50% spacing, TR 2300 ms, TE 2.98 ms).

Behavioral Analysis

For each participant and for each trial condition, mean reaction times (RTs) relative to the onset of S2, and response accuracies were calculated. Only correct responses were considered for further analyses. Haptic trials in which participants did not palpate the object, dropped the object, or made premature or late palpations as well as palpations lasting longer than 2 s were excluded from analysis (M ± SD: 0.06 ± 0.08 trials). Within each participant and condition, RTs that differed ±3 SDs from the preliminary mean were defined as outliers and excluded from further analyses (0.24 ± 0.18 trials). Mean RTs of the adjusted data were entered into a repeated-measures analysis of variance (PASW Statistics 18, SPSS Inc., Chicago, IL) with the factors S2-Modality (auditory/haptic), Congruency (congruent/incongruent), and Sensory-Matching (unimodal/crossmodal) as within-subject factors. Statistical effects at $P < 0.05$ were considered significant. Post hoc Bonferroni-corrected paired t-tests were used to test for differences between single conditions.

Image Analysis

Image processing and statistical analyses were performed using statistical parametric mapping 8 (SPM8; www.fil.ion.ucl.ac.uk/spm). The first 5 volumes of each time series were discarded to account for T_1 equilibrium effects. Data processing consisted of slice timing (correction for differences in slice acquisition time), realignment (rigid body motion correction), and unwarping (accounting for susceptibility by movement interactions), spatial normalization to Montreal Neurological Institute standard space as implemented in SPM8, thereby resampling to a voxel size of 3 × 3 × 3 mm^3, and smoothing with an 8 mm full-width at half-maximum isotropic Gaussian kernel. Statistical analysis was performed separately for each voxel using a general linear model.

By jittering the interval between S1 and S2, we were able to model the effects of object encoding (S1) and object matching (S2) independently. However, the main interest of this study was the interdependence of crossmodal matching and semantic congruency. Because matching of the object features was triggered by S2 presentation, our analysis focused on task-related changes in the BOLD signal elicited by S2. For brain activity related to auditory and haptic sample encoding (effects explained by the onset of S1) and the delay period between S1 and S2, please refer to Supplementary Methods and Results (Supplementary Fig. S1 and Table S2).

At the individual level (fixed effects), we defined separate regressors for the onset of S2 within each trial condition. Only correct trials withstanding the same inclusion criteria as applied for RT analyses were modeled. All onset vectors were modeled by convolving delta functions with a canonical hemodynamic response function and their first derivative. Low-frequency drifts in the BOLD signal were removed by a high-pass filter with a cutoff period of 128 s. We refrained from modeling the whole trials as compound events as the processing of S1 and the delay period until the S2 instruction tone (i.e., on average about 75% of the length of the trials) was noninformative in relation to the S2-matching condition.

The design matrix for statistical inferences (random effects) was configured using a "flexible factorial design" as implemented in SPM8 (Henson and Penny 2005). The model included the main effect of a Subject factor (modeling the participants' constants) and the interaction between our experimental factors. Thus, according to the 2 × 2 × 2 experimental design (S2-Modality × Congruency × Sensory-Matching), 8 S1–S2 condition regressors were estimated (A = auditory, H = haptic, c = congruent, i = incongruent): AAc, HAc, AAi, HAi, HHc, AHc, HHi, and AHi. We corrected for possible nonsphericity of the error term (dependence and possible unequal variance between conditions in the within-subject factors).

Main effect of interest was the interaction between the effect of Sensory-Matching (crossmodal > unimodal) and Congruency (congruent > incongruent), in the sense that we aimed to identify brain regions that were modulated by crossmodal matching only if the object pairs were semantically congruent but not if they were incongruent. We calculated these effects separately for auditory and haptic S2 conditions. In a next step, we investigated specific and general effects of S2-Modality (auditory vs. haptic) on this interaction. To this end, we used contrasts which compared modality-specific differential effects across auditory and haptic S2 conditions rather than directly comparing auditory versus haptic S2 processing. This procedure eliminated modality-specific confounding influences such as residual effects of the cue on S2 processing or motor activations during haptic exploration. Furthermore, as each object was presented 16 times throughout the experiment (8 times as auditory stimulus, 8 times as haptic stimulus), we conducted a control analysis to assess the potential influence of repeated object presentations on our results (see Supplementary Methods and Fig. S3).

If not stated otherwise, the reported regions were positively activated within the condition of interest in relation to baseline (inclusive masks at $P < 0.05$, uncorrected). It is important to note that the used baseline cannot be interpreted as a fixation baseline condition because we had not included null-events in our design. A description of auditory and haptic processing of the object stimuli used in this study in relation to fixation baseline can be found in Kassuba et al. (2011).

In this recent study, we have shown that the left pSTS and FG are both consistently stronger activated during the processing of auditory as well as haptic object stimuli relative to nonobject control stimuli (Kassuba et al. 2011). Therefore, we report voxelwise familywise error rate (FWE) corrected P values as obtained from small volume correction in these regions of interest ($P < 0.05$). Correction was based on spheres centered over peak coordinates obtained from the conjunction of auditory and haptic object-specific processing in our previous study (Kassuba et al. 2011). Based on the size of the regions identified in the previous study, the radius of the spheres was set to 8 mm: The FG was corrected using a sphere centered at $x = -39$, $y = -45$, $z = -18$ and the pSTS was corrected using a sphere centered at $x = -51$, $y = -57$, $z = 12$.

For all other voxels in the brain, voxelwise whole-brain FWE correction was applied ($P < 0.05$). Anatomical labeling was done by using the probabilistic stereotaxic cytoarchitectonic atlas implemented in the Anatomy Toolbox version 1.7 (Eickhoff et al. 2005) and adjusted according to anatomical landmarks of the average structural T_1-weighted image of all participants. The percent signal change plotted for visualization of the results was extracted using the rfxplot toolbox (Gläscher 2009).

Results

Task Performance

All participants solved the task with high accuracy. Mean frequency of correct responses over all conditions was 96.97 ± 2.63% with no significant differences between conditions ($P > 0.05$). Response latencies are displayed in Figure 2. Participants overall responded faster to auditory than haptic S2 (main effect of S2-Modality: $F_{1,20} = 238.89$, $P < 0.001$). Both, differences in RTs related to sensory matching as well as related to semantic congruency, were modulated by the modality in which S2 was processed (S2-Modality × Sensory-Matching interaction: $F_{1,20} = 14.21$, $P < 0.01$; S2-Modality × Congruency interaction: $F_{1,20} = 15.53$, $P < 0.001$).

We examined these interactions further by exploring the effects separately for each S2 modality. For auditory S2, faster responses were noted for unimodal as compared to crossmodal matching (main effect of Sensory-Matching for auditory S2: $F_{1,20} = 60.96$, $P < 0.001$), and this effect was larger for congruent than incongruent stimulus pairs (Sensory-Matching × Congruency interaction for auditory S2: $F_{1,20} = 11.66$, $P < 0.01$). In trials with

haptic S2, participants responded consistently faster to congruent than incongruent object pairs (main effect of Congruency for haptic S2: $F_{1,20} = 39.58$, $P < 0.001$). In addition, even though not significant when corrected for post hoc testing, participants tended to react faster to unimodal than crossmodal matching, but only when S1 and S2 were semantically congruent ($t_{20} = 2.39$, $P = 0.027$, uncorrected; see Fig. 2). Accordingly, there was a significant Sensory-Matching by Congruency interaction for haptic S2 ($F_{1,20} = 4.65$, $P < 0.05$).

Functional MRI

Crossmodal Matching by Semantic Congruency Interactions

As outlined in the introduction, we defined the left FG and pSTS as candidate regions where semantically congruent object features are unified across audition and touch. According to our hypothesis, we expected that the left FG and pSTS would show an increase in activation when semantically congruent auditory and haptic information had to be matched relative to unimodal matching. Moreover, we expected this crossmodal matching effect to occur only when both objects were semantically congruent, as only congruent information can be unified to a coherent representation. Analysis of task-related activation in response to S2 revealed an involvement of the left FG and the left pSTS in crossmodal matching of semantically congruent object features. In both cortical regions, we found the expected interaction between crossmodal matching and semantic congruency (crossmodal > unimodal × congruent > incongruent), but only in the left FG, this interaction was independent of the direction of crossmodal matching (see Table 1 and Figs 3 and 4).

In the left FG, regional activity increased during crossmodal matching of auditory and haptic object features, but only if auditory and haptic stimuli referred to the same object. Thus, the left FG showed a stronger BOLD response in semantically congruent but not in semantically incongruent conditions if an auditory S2 had to be matched with a haptic S1 (HA trials) or vice versa (AH trials) as compared to when S2 and S1 were presented within the same modality (AA and HH trials, respectively; peak x, y, z in millimeters: −39, −43, −23; $t_{140} = 3.93$, $P < 0.01$, corrected). Most importantly, a direct comparison of the crossmodal matching effect within congruent and incongruent trials (i.e., Sensory-Matching × Congruency interaction) revealed significantly greater BOLD changes in the left FG in congruent trials. This was indicated by a conjunction analysis which considered the Sensory-Matching by Congruency interaction term for auditory and haptic S2 (see Table 1). A possible concern may be that the interaction pattern was driven by the fact that in the unimodal congruent condition, the 2 objects were identical, whereas they were different (although conceptually matching) in the crossmodal congruent trials. Thus, a crossmodal matching effect that only exists in congruent but not incongruent trials may not be surprising. However, the FG interaction patterns were not only defined by a relative increase during crossmodal as compared to unimodal congruent matching but in addition by a relative increase during crossmodal congruent as compared to incongruent matching (see Fig. 3). Planned t-contrasts confirmed the observed interaction patterns (auditory S2: HAc > HAi: $t_{140} = 1.89$; haptic S2: AHc > AHi: $t_{140} = 1.42$; both $P < 0.05$, uncorrected).

In contrast to the left FG, the left pSTS was only involved in matching an auditory S2 to a haptic S1 (HA trials) but not vice versa (AH trials; see Fig. 4). The left pSTS and adjacent pSTG showed a stronger activation when an auditory S2 had to be matched to a congruent haptic S1 as compared to unimodal auditory–auditory matching (−54, −52, 13; $t_{140} = 4.08$, $P < 0.01$, corrected). Critically, the increased activation during crossmodal as opposed to unimodal matching of auditory S2 was significantly greater in semantically congruent relative to

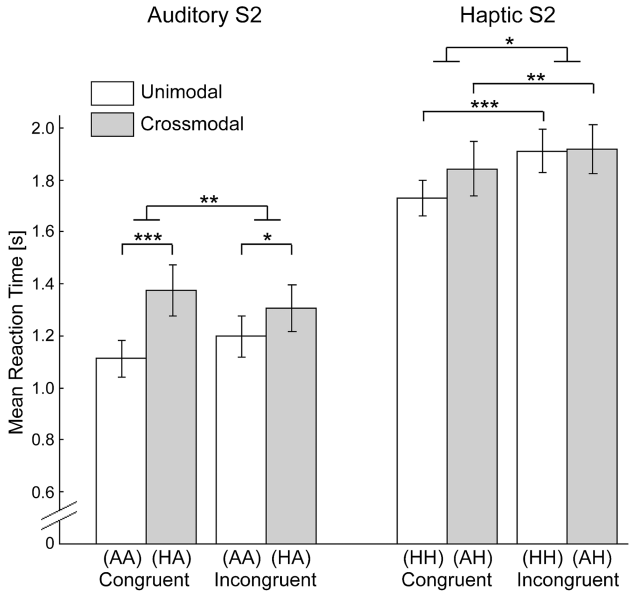

Figure 2. Mean RTs for auditory and haptic S2 for all 4 matching conditions (unimodal/crossmodal × congruent/incongruent) with error bars indicating the standard error of the mean. RTs were recorded from S2 onset onward. Responses to auditory S2 were significantly faster when they were matched with an auditory S1 (unimodal) compared to a haptic S1 (crossmodal). On the other hand, participants responded faster to haptic S2 if they referred to the same object concept as the S1 (congruent) than when they did not (incongruent). In general, it took participants longer to match congruent stimuli across than within modalities, and this difference was less (auditory S2) or not at all (haptic S2) pronounced for incongruent stimuli (sensory matching by semantic congruency interaction). Sample–target (S1–S2) conditions: A = auditory, H = haptic. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, Bonferroni corrected.

Table 1

Crossmodal matching by semantic congruency interaction (crossmodal > unimodal × congruent > incongruent)

Region	x	y	z	t_{peak}	$P_{corrected}$
Auditory and haptic S2 conditions: [(HAc > AAc) > (HAi > AAi)] ∩ [(AHc > HHc) > (AHi > HHi)]					
L FG	−39	−37	−20	3.08	0.026[a]
Only auditory S2 conditions: (HAc > AAc) > (HAi > AAi)					
L pSTS	−48	−52	7	3.16	0.021[b]
L/R medial frontal gyrus	6	11	52	5.44	0.004
L anterior cingulum	−6	23	31	4.70	0.067
Only haptic S2 conditions: (AHc > HHc) > (AHi > HHi)					
L LO	−48	−64	−5	5.00	0.022

Note: Coordinates are denoted by x, y, z in millimeters (Montreal Neurological Institute [MNI] space) and indicate the peak voxel. Strength of activation is expressed in t and P values corrected for the whole brain (df = 140). All contrasts are masked inclusively with the contrast of the respective condition of interest versus baseline ($P < 0.05$, uncorrected). There were no significant results for the reversed interaction contrasts (crossmodal > unimodal × incongruent > congruent) that surpassed a correction for multiple comparisons. L = left, R = right. Sample–target (S1-S2) conditions: A = auditory, H = haptic, c = congruent, i = incongruent.

[a]Corrected for the left FG.

[b]Corrected for the left pSTS.

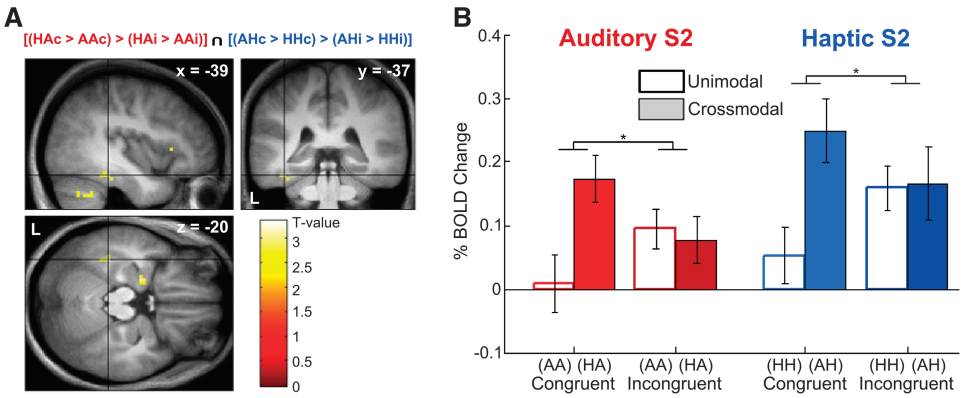

Figure 3. Activation pattern in the left FG during auditory-to-haptic and haptic-to-auditory matching of semantically congruent object features. (*A*) SPMs showing a stronger increase in activation of left FG to crossmodal relative to unimodal matching for congruent as opposed to incongruent object pairs (crossmodal > unimodal × congruent > incongruent). Increased activation during crossmodal matching of semantically congruent object features was found for both directions of crossmodal matching (HA and AH trials). For illustrative purposes, the statistical maps are thresholded at $P < 0.01$, uncorrected, and overlaid on the average structural T_1-weighted image of all participants. (*B*) Percent signal change and error bars indicating the standard error of the mean at the left FG (Montreal Neurological Institute [MNI] coordinates: $x, y, z = -39, -37, -20$) for each condition. Only in semantically congruent but not incongruent conditions, the left FG showed a stronger activation during crossmodal (HA and AH trials, respectively) relative to unimodal (AA and HH trials, respectively) matching. Moreover, activation was relatively decreased during unimodal congruent matching and relatively increased during crossmodal congruent matching as compared to incongruent matching. L = left. Sample–target (S1–S2) conditions: A = auditory, H = haptic, c = congruent, i = incongruent. *$P < 0.05$, corrected.

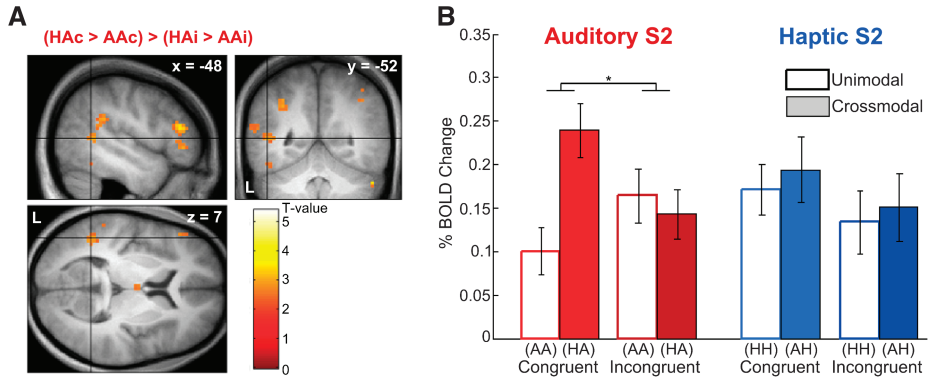

Figure 4. Activation pattern in the left pSTS during crossmodal matching of auditory S2 to semantically congruent S1. (*A*) For auditory S2, the increase in BOLD response in left pSTS to crossmodal compared to unimodal matching was greater for congruent than incongruent object pairs (crossmodal > unimodal × congruent > incongruent). For illustrative purposes, the statistical maps are thresholded at $P < 0.01$, uncorrected, and overlaid on the average structural T_1-weighted image of all participants. (*B*) Percent signal change and error bars indicating the standard error of the mean at the left pSTS (Montreal Neurological Institute [MNI] coordinates: $x, y, z = -48, -52, 7$) for each condition. Only in semantically congruent but not incongruent conditions, the left pSTS showed a stronger activation if an auditory S2 had to be matched to a haptic S1 as compared to when S2 and S1 were both auditory (red bars). Moreover, activation was relatively decreased during unimodal congruent matching and relatively increased during crossmodal congruent matching as compared to incongruent matching of auditory S2. This interaction effect did not occur for the analog comparisons when haptic S2 were matched (blue bars). L = left. Sample–target (S1–S2) conditions: A = auditory, H = haptic, c = congruent, i = incongruent. *$P < 0.05$, corrected.

incongruent trials (Sensory-Matching × Congruency interaction within auditory S2; see Table 1). Importantly, the activation was relatively increased during crossmodal congruent matching as compared to incongruent matching (planned *t*-contrast: HAc > HAi: $t_{140} = 2.03$, $P < 0.05$, uncorrected; see Fig. 4). In contrast, even though the pSTS was consistently activated when processing haptic S2, the interaction of crossmodal matching and semantic congruency was not significant for haptic S2 ($P > 0.7$, uncorrected). The interaction effect found for auditory S2 trended to be significantly greater than for haptic S2 (Sensory-Matching × Congruency × S2-Modality; –48, –52, 7; $t_{140} = 2.56$, $P = 0.09$, corrected).

A whole-brain analysis revealed a bilateral increase in activation in the LO when a haptic S2 was matched to a congruent auditory S1 as compared to when S2 and S1 were both presented haptically (left LO: –48, –64, –8; $t_{140} = 8.96$; right LO: 48, –58, –11, $t_{140} = 7.00$; both $P < 0.001$, corrected). In contrast, the activation in the LO did not differ between unimodal and crossmodal matching if S1 and S2 were semantically incongruent. Accordingly, the crossmodal matching effect (crossmodal > unimodal) in the left LO was significantly greater for semantically congruent than incongruent stimulus pairs (Sensory-Matching × Congruency within haptic S2; see Table 1 and Fig. 5). However, in contrast to the interaction patterns found in the FG and pSTS, the activation in the LO did not differ between crossmodal matching of congruent versus incongruent input. The left LO was rather strongly activated whenever a haptic S2 was processed and, thereby, less activated when the same haptic input was matched (see Fig. 5). For auditory S2, on the other hand, there was no interaction effect at all ($P > 0.5$, uncorrected). In fact, the left LO showed rather a decrease in the BOLD response in all conditions in which an auditory S2 was processed (see Fig. 5).

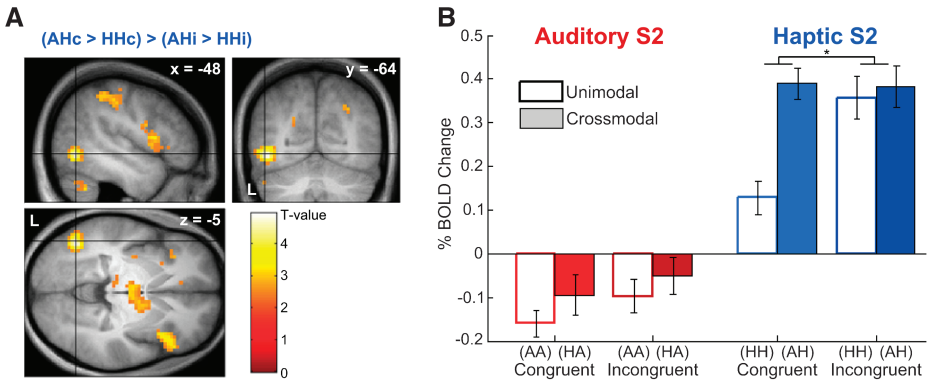

Figure 5. Activation pattern in the left LO during crossmodal matching of haptic S2 to semantically congruent S1. (*A*) For haptic S2, the increase in BOLD response in left LO to crossmodal compared to unimodal matching was greater for congruent than incongruent object pairs (crossmodal > unimodal × congruent > incongruent). For illustrative purposes, the statistical maps are thresholded at $P < 0.01$, uncorrected, and overlaid on the average structural T_1-weighted image of all participants. (*B*) Percent signal change and error bars indicating the standard error of the mean at the left LO (Montreal Neurological Institute [MNI] coordinates: $x, y, z = -48, -64, -5$) for each condition. Only in semantically congruent but not incongruent conditions, the left LO showed a stronger activation if a haptic S2 had to be matched to an auditory S1 as compared to when S2 and S1 were both haptic (blue bars). This interaction effect did not occur for the analog comparisons when auditory S2 were matched (red bars). The left LO was rather deactivated in all auditory S2 conditions. L = left. Sample–target (S1–S2) conditions: A = auditory, H = haptic, c = congruent, i = incongruent. *$P < 0.05$, corrected.

The reversed interaction contrast (crossmodal > unimodal × incongruent > congruent) did not yield any significant results in any of the S2-modality conditions. Taken together, the left FG showed a crossmodal matching effect that was specific to semantically congruent object features and independent of the direction of matching. In contrast, the left pSTS displayed the interaction only for auditory and the left LO only for haptic S2. Most important, this selectivity to congruent crossmodal matching cannot be explained by task difficulty because the RT data indicated that in general, crossmodal incongruent matching was at least as difficult or more difficult than crossmodal congruent matching.

Main Effects of Crossmodal Matching

In addition to the effects modulated by semantic congruency, we also found crossmodal matching effects that were independent of semantic congruency (for a summary, see Table 2). The left inferior frontal gyrus (IFG) and a region in the left posterior FG (pFG; posterior to the region reported above) were both stronger activated during crossmodal than during unimodal matching for both auditory and haptic S2 and independent of semantic congruency. In addition, modality-specific main effects for crossmodal matching were found in secondary sensory cortices and association cortices such as in bilateral pSTG for auditory S2 and bilateral precentral and postcentral gyri as well as the anterior insula for haptic S2. The reversed contrast (unimodal > crossmodal) did not yield any significant results.

Main Effects of Semantic Congruency

Effects of semantic congruency or incongruency that were not modulated by crossmodal matching were only found for haptic S2 (see Table 3). A region in left pFG and the IFG displayed an incongruency effect (i.e., stronger activation for incongruent vs. congruent matching) for haptic S2, independent of S1-modality. A reversed effect (congruent > incongruent) was found in the left angular gyrus, but only when the threshold was lowered to $P < 0.001$, uncorrected.

The activations found in the left IFG and pFG for the incongruency contrast partially overlapped with the activations found for the main effect of crossmodal matching contrast (see Supplementary Figs S5 and S6). Thus, the 2 regions showed increased BOLD responses in S2-matching conditions in which participants needed longer to react (crossmodal matching for auditory S2, incongruent matching for haptic S2), suggesting that the activation of these regions might have been driven by task difficulty. An additional analysis supported this assumption by showing that the individual trialwise BOLD responses in these regions (but not in the more anterior mid-FG region showing the crossmodal matching by semantic congruency interaction effect) were positively correlated with individual trialwise response latencies (see Supplementary Methods, Results, and Figs S5 and S6).

Discussion

Here, we performed fMRI during a delayed-match-to-sample task to investigate audio-haptic interactions during higher order object recognition. In secondary auditory and somatosensory cortices and association cortices, the BOLD response increased during crossmodal as opposed to unimodal object matching independently of semantic congruency. In contrast, the left FG and pSTS were specifically activated during crossmodal matching of a target (S2) with a congruent sample (S1). The direction of the observed activity patterns suggests complementary functions of these regions: Only the left FG showed a stronger differential activation during both haptic-to-auditory and auditory-to-haptic matching of semantic congruent object features. By contrast, the left pSTS was selectively engaged in matching auditory targets. The present results extend previous findings that the left FG and pSTS are object selective within the auditory and tactile modalities (Kassuba et al. 2011). Together, the results show that auditory and haptic object processing interact at various stages of object recognition and putatively converge into a higher order conceptual object representation hosted by the FG.

Audio-haptic Interactions in the FG

The left FG showed a selective crossmodal matching effect for congruent but not incongruent trials. Moreover, there was an additional increase in activation during crossmodal congruent as compared to crossmodal incongruent matching. As

Table 2
Main effects crossmodal matching (crossmodal vs. unimodal)

Region	*x*	*y*	*z*	t_{peak}	$P_{corrected}$
Crossmodal > unimodal, auditory and haptic S2 conditions: (HAc > AAc) ∩ (HAi > AAi) ∩ (AHc > HHc) ∩ (AHi > HHi)					
L pFG	−42	−49	−11	3.43	0.013[a]
L IFG (pars triangularis)	−45	14	25	4.64	0.082
Crossmodal > unimodal, auditory S2 conditions: (HAc > AAc) ∩ (HAi > AAi)					
L pFG	−42	−49	−11	3.70	0.006[a]
L IFG (pars triangularis)	−42	29	7	5.91	0.001
L IFG (pars triangularis)	−42	11	28	4.93	0.029
L superior temporal gyrus	−42	−28	7	5.34	0.006
L superior temporal gyrus	−51	−19	4	4.95	0.027
R superior temporal gyrus	45	−22	4	6.03	<0.001
R superior temporal gyrus	51	−10	1	5.92	0.001
R superior temporal gyrus	57	−4	−8	4.99	0.023
L cerebellum	−33	−46	−29	4.89	0.033
Crossmodal > unimodal, haptic S2 conditions: (AHc > HHc) ∩ (AHi > HHi)					
L pFG	−42	−49	−17	4.20	0.001[a]
L IFG (pars opercularis)	−48	11	28	5.13	0.014
L IFG (pars triangularis)	−39	20	22	4.87	0.036
L IFG (pars triangularis)	−48	29	16	4.85	0.039
L IFG (pars orbitalis)	−27	32	−2	4.89	0.033
L/R middle cingulum	0	−7	52	6.92	<0.001
L middle cingulum	−3	2	40	6.32	<0.001
L anterior cingulum	−6	23	31	6.48	<0.001
L superior medial gyrus	−9	38	28	4.80	0.047
L precentral gyrus	−30	−16	55	5.83	0.001
L postcentral gyrus	−42	−19	49	6.21	<0.001
L postcentral gyrus	−30	−34	55	5.66	0.002
R precentral gyrus	36	−19	64	4.84	0.041
R postcentral gyrus	33	−34	55	5.33	0.006
R postcentral gyrus	54	−13	37	5.08	0.016
L anterior insula	−24	29	1	4.79	0.049
R anterior insula	30	29	−2	5.28	0.008
L thalamus	−12	−19	7	5.41	0.004

Note: Coordinates are denoted by *x*, *y*, *z* in millimeters (Montreal Neurological Institute [MNI] space) and indicate the peak voxel. Strength of activation is expressed in *t* and *P* values corrected for the whole brain (df = 140). All contrasts are masked inclusively with the contrast of the respective condition of interest versus baseline ($P < 0.05$, uncorrected). There were no significant results for the respective reversed contrasts (unimodal > crossmodal) that surpassed a correction for multiple comparisons. L = left, R = right. Sample−target (S1–S2) conditions: A = auditory, H = haptic, c = congruent, i = incongruent.
[a]Corrected for the left FG.

Table 3
Main effects semantic congruency (congruent vs. incongruent)

Region	*x*	*y*	*z*	t_{peak}	$P_{corrected}$ ($P_{uncorrected}$)
Congruent > incongruent, haptic S2 conditions: (AHc > AHi) ∩ (HHc > HHi)					
L angular gyrus	−51	−58	37	4.07	(<0.001)
Incongruent > congruent, haptic S2 conditions: (AHi > AHc) ∩ (HHi > HHc)					
L pFG	−42	−52	−14	3.31	0.019[a]
L IFG (pars opercularis)	−42	8	25	4.90	0.032
L IFG (pars triangularis)	−42	29	22	4.65	0.080
L superior medial gyrus	−6	17	46	4.52	(<0.001)
R IFG (pars triangularis)	51	20	28	4.13	(<0.001)
R IFG (pars triangularis)	45	32	16	3.45	(<0.001)
R middle frontal gyrus	42	29	25	3.55	(<0.001)
L middle OccG/pIPS	−30	−70	37	3.88	(<0.001)

Note: Coordinates are denoted by *x*, *y*, *z* in millimeters (Montreal Neurological Institute [MNI] space) and indicate the peak voxel. Strength of activation is expressed in *t* and *P* values corrected for the whole brain (df = 140). All contrasts are masked inclusively with the contrast of the respective condition of interest versus baseline ($P < 0.05$, uncorrected). Additional clusters at $P < 0.001$, uncorrected, are listed if they exceed a cluster size > 15 voxel (*P* values in parentheses). There were no significant results for the respective contrasts in auditory S2 conditions that surpassed the above used thresholds. L = left, R = right. OccG = occipital gyrus, pIPS = posterior intraparietal sulcus. Sample−target (S1–S2) conditions: A = auditory, H = haptic, c = congruent, i = incongruent.
[a]Corrected for the left FG.

coherence in content is a relevant factor for the binding of meaningful information (Laurienti et al. 2004), we speculate that the specific neuronal enhancement for congruent crossmodal matching might imply multisensory binding mechanisms related to a coherence in meaning. The current results extend previous data showing that the FG processes object stimuli across vision and touch (e.g., James et al. 2002; Stevenson et al. 2009), vision and audition (e.g., Adams and Janata 2002; Naumer et al. 2009; Stevenson and James 2009), or all 3 modalities (Binkofski et al. 2004; Kassuba et al. 2011) and suggest that the FG is a higher order convergence zone for multisensory object information.

Recognizing an object by a related sound is arbitrary and depends on prior associations of a given object and the respective sound (Griffiths and Warren 2004). Accordingly, in order to match object information across audition and touch, associative object knowledge needs to be accessed. Together with other studies of semantic processing (Wagner et al. 1998; Price 2000; Wheatley et al. 2005; Gold et al. 2006), the current data indicate a contribution of the FG when detailed semantic interpretations are involved such a linking the typical sound, shape, and consistency associated with a given object. In line with this view, learning audio-visual associations between artificial objects results in an integration-related recruitment of the left FG (Naumer et al. 2009). Therefore, the left FG might provide conceptual object representations (Martin 2007) in the form of associative object knowledge that is accessible via different senses.

Given the current data, we can only speculate about how the FG convergence zone might bind information from different senses at the level of single neurons. Whereas the gold standard for multisensory integration advancing the orientation of behavior is neuronal convergence (i.e., synaptic convergence of modality-specific processing onto individual neurons; Meredith 2002), much less is known about the physiological principles underlying higher order processes such as perceptual binding (Stein and Stanford 2008). Theories on semantic memory suggest that multisensory associative object knowledge is implemented as supramodal nodes linking features provided by different senses (Damasio 1989; Mesulam 1998; Meyer and Damasio 2009). Thereby, the neurons within associative areas code a record to reconstruct an approximation of the original perceptual representations. At the level of single neurons, such records might be formed by strengthened synaptic interconnections due to Hebbian learning (Cruikshank and Weinberger 1996) between neurons preferentially connected to specific unisensory areas or lower order convergence areas (Mesulam 1998; Meyer and Damasio 2009). Given the role of the FG in processing meaning (Martin and Chao 2001; Martin 2007), it is tempting to speculate that the FG directly links semantic attributes related to manipulable objects (e.g., graspable shape, nonbiological motion pattern, related action) extracted by lower order convergence zones rather than perceptual features per se.

An alternative account is that the activation of the left FG might be related to visual imagery (Ishai et al. 2000; O'Craven and Kanwisher 2000), used by the participants as strategy to accomplish task performance. Although, visual imagery might have occurred to some degree, 2 reasons render it unlikely that it explains the differential activation pattern in the FG. First, if auditory and haptic object input were translated into visual images, one might expect the visual images to be particularly maintained during the delay period. However, there was no activation in the FG nor in any other visual region during the delay period (see Supplementary Table S2). The fact that in

contrast the left IFG showed a sustained activation during the delay period might suggest that participants encoded the samples semantically (e.g., as object names) rather than visually. Second, if the differential activation pattern in the left FG at time point of matching was driven by differences in visual imagery, we would expect to find a similar differential activation pattern in the left LO which had shown a similar object-specific activation pattern as the left FG in a previous study using the visual counterparts of the currently used object stimuli (Kassuba et al. 2011; see also Lacey et al. 2010). The left LO, however, showed no clear selectivity for congruent over incongruent crossmodal matching and was rather deactivated during auditory target conditions.

Audio-haptic Interactions in the pSTS

The left pSTS only showed a greater crossmodal matching effect for congruent as opposed to incongruent matching when auditory but not when haptic targets had to be processed. This observation extends our previous findings that the left pSTS is increasingly activated during processing of action-related object sounds as well as during active palpation of the respective manipulable objects as compared to non-object control stimuli (Kassuba et al. 2011). The current results tie in with previous work in nonhuman primates (Kayser et al. 2005) and humans (Foxe et al. 2002; Beauchamp et al. 2008) showing that multisensory interactions between low-level auditory and tactile input in the pSTG sometimes extend into the pSTS, as well as with human EEG data source reconstruction suggesting that haptic-to-auditory object priming effects are located in the left lateral temporal cortex (Schneider et al. 2011). The pSTG and adjacent pSTS have been particularly implicated in higher order auditory functions such as storing auditory knowledge related to objects (Wheeler et al. 2000; James and Gauthier 2003) and making semantic judgments on those auditory object features (Kellenbach et al. 2001; Goldberg et al. 2006). We, therefore, argue that conceptual object representations in left pSTS are primarily auditory in nature but can be activated in a top-down fashion by crossmodal feedback loops.

Previous work has implicated the pSTS in audio-visual integration of meaningful information (for reviews, see Amedi et al. 2005; Doehrmann and Naumer 2008). For instance, Tanabe et al. (2005) found a similar asymmetry of the activation in left pSTS in a delayed-match-to-sample learning task between the matching of visual and auditory stimuli. More posterior parts in left pSTS showed a decrease in activation when learning audio-visual associations when auditory targets were matched to visual samples but not vice versa (see Fig. 5*A*/*B* and Table 2 in Tanabe et al. 2005). Similarly, a region in left pSTG and adjacent pSTS/middle temporal gyrus was selectively modulated by auditory but not by visual or action-related conceptual features, suggesting that this region codes higher order acoustic object information contributing to a concept (Kiefer et al. 2008). Together with the current results, these previous data support the assumption that the left pSTS is primarily tuned to auditory object information. It has to be kept in mind that matching of an auditory target with a haptic sample is constantly taking place in daily life since it is usually the manipulation of the object that produces a characteristic sound. Thus, the interactions of auditory and haptic object features in left pSTS might be strongly influenced by the pragmatic use of these manipulable objects in daily life.

Audio-haptic Interactions in the LO

The left LO might be the counterpart of the pSTS hosting haptic object representations that are crossmodally modulated by sounds because the left LO displayed the opposite interaction, namely a selective increase in activity during crossmodal matching of a haptic target to a congruent auditory sample. In contrast to the left pSTS, the LO was only activated when processing haptic but not auditory targets, which ties in with previous studies (e.g., Amedi et al. 2001, 2002; Kassuba et al. 2011). This may be explained by the particular involvement of the LO in processing the shape of objects, a feature which is normally derived from visual and haptic but not auditory input (Amedi et al. 2002, 2007). In a recent fMRI-adaptation (fMRI-A) study, Doehrmann et al. (2010) found an enhanced response to auditory object repetitions in contrast to auditory object changes in the left LO, thus, demonstrating that the object-related left LO is able to discriminate auditory object stimuli at least along the dimension "same" versus "different." Together with the crossmodal matching by semantic congruency interaction for haptic targets found in the current results, we speculate that the LO is able to extract conceptual object information from auditory input to a certain degree without being an auditory region per se (cf. Amedi et al. 2007; Kim and Zatorre 2011).

Audio-haptic Interactions at Different Hierarchical Stages of Object Processing

Regardless of semantic congruency, a wide range of regions along the proposed auditory (Ahveninen et al. 2006) and haptic (Reed et al. 2005) "what" pathways displayed an increased activation during crossmodal as compared to unimodal matching of auditory and haptic targets. For instance, crossmodal matching of auditory targets involved the bilateral pSTG, whereas crossmodal matching of haptic targets activated the bilateral anterior insula as well as the bilateral precentral and postcentral gyri. The pSTG, a part of the auditory association cortex (Kaas and Hackett 2000) and the possible homologous area in the caudal auditory belt in monkeys, has previously been shown to integrate low-level auditory and somatosensory information in humans (Foxe et al. 2002; Schurmann et al. 2006) and monkeys (Kayser et al. 2005), respectively. Similarly, auditory and somatosensory "what" information have been shown to converge in the bilateral anterior insula (Renier et al. 2009). Thus, the effects for crossmodal matching independent of semantic congruency found in these regions in the current data might very likely reflect early audio-haptic interactions that are not related to object recognition per se. However, the crossmodal interactions in auditory-related cortices were clearly biased toward auditory object processing, whereas the crossmodal interactions in somatosensory-related cortices were biased toward haptic object processing.

Modality-independent crossmodal matching effects that were not affected by semantic congruency were found in the left IFG and a left pFG region, just posterior to the region where crossmodal matching was modulated by semantic congruency. The same regions were also stronger activated during matching of incongruent as opposed to congruent haptic targets. Similar incongruency effects have been found in the left IFG for audio-visual interactions (Hein et al. 2007; Noppeney et al. 2008). These findings imply that in the context of the current task, the left IFG might have played the role of a

semantic working memory or executive system (Poldrack et al. 1999), possibly involved in semantic conflict monitoring and modulated by task difficulty. In line with this interpretation is also that the left IFG showed a sustained activation during the delay period between sample and target and that its activity was in general positively correlated with individual response latencies (see Supplementary Results, Table S2, and Fig. S5). The left pFG region might similarly be involved in higher order processes and probably converge auditory and haptic information, which is then unified into coherent object representations in the more anterior FG region (Martin 2007). Apart from the main effects for crossmodal matching, a main effect of semantic congruency that was independent of crossmodal matching was only found as a trend for haptic targets in the left angular gyrus. This trend might reflect the involvement of the angular gyrus in semantic retrieval (Noppeney and Price 2003, 2004; Noppeney et al. 2008).

Together, the results suggest multisensory interactions between audition and touch at various stages of object recognition: Object processing in secondary sensory cortices and association cortices is crossmodally modulated independent of semantic congruency, whereas crossmodal interactions in the left FG and the left pSTS are highly dependent on semantic congruency between object pairs, indicating that they unify crossmodal information into a coherent object representation. The left pSTS might host auditory-dominated object concepts that can be crossmodally modulated. Conversely, the left FG might represent a higher order convergence zone for object concepts independent of the modality the object is perceived in.

Methodological Considerations

In addition to more basic binding cues like temporal and spatial coherence (Stein and Stanford 2008), the correspondence in content (semantic congruency) is an important factor for the binding of object features provided by different senses into a unified conceptual representation (Laurienti et al. 2004). However, unlike most previous studies comparing the effects of semantic congruency during simultaneous processing of input from different senses (for review, see Doehrmann and Naumer 2008), we introduced a temporal delay between the processing of the 2 object stimuli. Consequently, the study focused on the effects of audio-haptic interactions on higher order object recognition rather than basic multisensory integration processes (Stein and Stanford 2008). Our choice was based on the following considerations: simultaneous processing of 2 auditory stimuli potentially reduces their perceptibility and the concurrent haptic exploration of 2 objects would have been hardly possible with one hand. By employing a delayed-match-to-sample paradigm, we kept the perceptual requirements constant across the unimodal and crossmodal conditions. Moreover, the use of a semantic task guaranteed a conceptual processing of both target and sample.

The sequential stimulus presentation employed in the present study resembles fMRI-A or priming paradigms. fMRI-A studies revealed a crossmodal adaptation effect for semantic congruent information in multisensory integration regions such as in the pSTS across audition and vision (van Atteveldt et al. 2010) and in the LO across vision and touch (Tal and Amedi 2009). In the current study, we rather found the opposite effect, namely increased activation during crossmodal congruent matching instead of an adaptation in object-specific audio-haptic regions of interest. Furthermore, we did not find any general habituation effects of the BOLD response during target matching due to the repeated presentation of the same object stimuli throughout the experiment (see Supplementary Results and Fig. S3). There are 2 major differences between the present study and typical fMRI-A paradigms that might account for this difference. First, the stimulus onset asynchronies between sample and target were rather long and might have favored a semantic encoding of the sample. Second, instead of using a task orthogonal to the effect of interest (crossmodal matching) such as a detection task (Doehrmann et al. 2010; van Atteveldt et al. 2010) or passive recognition (Tal and Amedi 2009), our task explicitly required a semantic decision on whether the 2 objects matched onto the same concept.

The exact neuronal mechanisms underlying BOLD adaptation are not readily understood and probably vary as a function of the time scale of adaptation, task and stimuli, and brain regions (Grill-Spector et al. 2006). Thus, the task demands in the current study might have overruled general effects of stimulus habituation. In line with this account, the BOLD response during sample encoding but not during target matching in the left pFG and pSTS habituated as a function of how often an object had already been presented throughout the experiment (see Supplementary Figs S4 vs. S3). Similarly in the visuo-haptic domain, for instance, the left anterior intraparietal sulcus showed crossmodal adaptation effects in an fMRI-A paradigm (Tal and Amedi 2009), whereas it showed enhanced responses to crossmodal matching in a delayed-match-to-sample task (Grefkes et al. 2002). Yet, our findings are consistent with the results from crossmodal fMRI-A studies (e.g., Tal and Amedi 2009; van Atteveldt et al. 2010), namely by showing a selectivity of occipitotemporal regions for crossmodal associations of meaningful object information.

Finally, the comparison of conditions with versus without haptic exploration differs in terms of manual activity, which is present during the former but not the latter (Naumer et al. 2010). Therefore, we refrained from directly comparing haptic versus auditory target conditions but rather limited the comparison to modality-specific differential effects across modalities. This enabled us to eliminate motor specific effects related to haptic target processing by the applied contrasts. While we cannot completely rule out that motor activity during sample encoding had affected target processing (HA > AA and HH > AH, respectively), this is unlikely for 2 reasons. First, the jittered delay of 2–6 s between sample and target objects allowed us to model BOLD responses to the target objects independently from BOLD responses to the sample objects. Second, if a differential pattern of hand movements due to haptic sample encoding had affected target processing, we would expect to consistently find increased activations in motor cortex for conditions with versus without haptic samples, which was not the case (see Supplementary Results).

Conclusions

In summary, the present fMRI study used a delayed-match-to-sample paradigm to study audio-haptic interactions in higher order object recognition. We propose that audition and touch interact at various stages of object processing with crossmodal modulations at early association cortices and intermediate modality-sensitive object recognition regions (pSTS and putatively LO) and converge into higher order conceptual object representations in the left FG. Future studies might shed more

light on the role of the FG in multisensory object recognition and particularly in the integration of auditory and haptic object information.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

Funding

This study was funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF; 01GW0562 to H.R.S.). T.K. was supported by the Janggen-Pöhn Stiftung and the Jubiläumsstiftung der Basellandschaftlichen Kantonalbank. M.M.M. was supported by the BMBF and by the FP7 project ROSSI, Emergence of communication in RObots through Sensorimotor and Social Interaction (216125). H.R.S. was supported by a Grant of Excellence on the control of actions "ContAct" from the Lundbeck Foundation (R59 A5399).

Notes

We thank Gesine Müller, Katrin Wendt, and Kathrin Müller for their support in acquiring the fMRI data, Jürgen Finsterbusch for setting up the fMRI sequence, and Kristoffer Hougaard Madsen for helpful suggestions on the data analysis. *Conflict of Interest* : None declared.

References

Adams RB, Janata P. 2002. A comparison of neural circuits underlying auditory and visual object categorization. Neuroimage. 16:361–377.

Ahveninen J, Jaaskelainen IP, Raij T, Bonmassar G, Devore S, Hamalainen M, Levanen S, Lin FH, Sams M, Shinn-Cunningham BG, et al. 2006. Task-modulated "what" and "where" pathways in human auditory cortex. Proc Natl Acad Sci U S A. 103:14608–14613.

Amedi A, Jacobson G, Hendler T, Malach R, Zohary E. 2002. Convergence of visual and tactile shape processing in the human lateral occipital complex. Cereb Cortex. 12:1202–1212.

Amedi A, Malach R, Hendler T, Peled S, Zohary E. 2001. Visuo-haptic object-related activation in the ventral visual pathway. Nat Neurosci. 4:324–330.

Amedi A, Stern WM, Camprodon JA, Bermpohl F, Merabet L, Rotman S, Hemond C, Meijer P, Pascual-Leone A. 2007. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. Nat Neurosci. 10:687–689.

Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ. 2005. Functional imaging of human crossmodal identification and object recognition. Exp Brain Res. 166:559–571.

Annett M. 1970. A classification of hand preference by association analysis. Br J Psychol. 61:303–321.

Beauchamp MS. 2005. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Curr Opin Neurobiol. 15:145–153.

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. 2004. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci. 7:1190–1192.

Beauchamp MS, Lee KE, Argall BD, Martin A. 2004. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron. 41:809–823.

Beauchamp MS, Yasar NE, Frye RE, Ro T. 2008. Touch, sound and vision in human superior temporal sulcus. Neuroimage. 41:1011–1020.

Binkofski F, Buccino G, Zilles K, Fink GR. 2004. Supramodal representation of objects and actions in the human inferior temporal and ventral premotor cortex. Cortex. 40:159–161.

Cruikshank SJ, Weinberger NM. 1996. Receptive-field plasticity in the adult auditory cortex induced by Hebbian covariance. J Neurosci. 16:861–875.

Damasio AR. 1989. Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. Cognition. 33:25–62.

Doehrmann O, Naumer MJ. 2008. Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. Brain Res. 1242:136–150.

Doehrmann O, Weigelt S, Altmann CF, Kaiser J, Naumer MJ. 2010. Audio-visual functional magnetic resonance imaging adaptation reveals multisensory integration effects in object-related sensory cortices. J Neurosci. 30:3370–3379.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage. 25:1325–1335.

Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W, Murray MM. 2002. Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J Neurophysiol. 88:540–543.

Gläscher J. 2009. Visualization of group inference data in functional neuroimaging. Neuroinformatics. 7:73–82.

Gold BT, Balota DA, Jones SJ, Powell DK, Smith CD, Andersen AH. 2006. Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. J Neurosci. 26:6523–6532.

Goldberg RF, Perfetti CA, Schneider W. 2006. Perceptual knowledge retrieval activates sensory brain regions. J Neurosci. 26:4917–4921.

Grefkes C, Weiss PH, Zilles K, Fink GR. 2002. Crossmodal processing of object features in human anterior intraparietal cortex: an fMRI study implies equivalencies between humans and monkeys. Neuron. 35:173–184.

Griffiths TD, Warren JD. 2004. What is an auditory object? Nat Rev Neurosci. 5:887–892.

Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci. 10:14–23.

Hein G, Doehrmann O, Muller NG, Kaiser J, Muckli L, Naumer MJ. 2007. Object familiarity and semantic congruency modulate responses in cortical audio-visual integration areas. J Neurosci. 27:7881–7887.

Henson R, Penny W. 2005. ANOVAs and SPM. Technical report. London: Wellcome Department of Imaging Neuroscience. 1–24. Available from: http://www.fil.ion.ucl.ac.uk/~wpenny/publications/rik_anova.pdf (last accessed 8 March, 2012).

Ishai A, Ungerleider LG, Haxby JV. 2000. Distributed neural systems for the generation of visual images. Neuron. 28:979–990.

James TW, Gauthier I. 2003. Auditory and action semantic features activate sensory-specific perceptual brain regions. Curr Biol. 13:1792–1796.

James TW, Humphrey GK, Gati JS, Servos P, Menon RS, Goodale MA. 2002. Haptic study of three-dimensional objects activates extrastriate visual areas. Neuropsychologia. 40:1706–1714.

Kaas JH, Hackett TA. 2000. Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci U S A. 97:11793–11799.

Kassuba T, Klinge C, Hölig C, Menz MM, Ptito M, Röder B, Siebner HR. 2011. The left fusiform gyrus hosts trisensory representations of manipulable objects. Neuroimage. 56:1566–1577.

Kayser C, Petkov CI, Augath M, Logothetis NK. 2005. Integration of touch and sound in auditory cortex. Neuron. 48:373–384.

Kellenbach ML, Brett M, Patterson K. 2001. Large, colorful, or noisy? Attribute- and modality-specific activations during retrieval of perceptual attribute knowledge. Cogn Affect Behav Neurosci. 1:207–221.

Kiefer M, Sim EJ, Herrnberger B, Grothe J, Hoenig K. 2008. The sound of concepts: four markers for a link between auditory and conceptual brain systems. J Neurosci. 28:12224–12230.

Kim J-K, Zatorre RJ. 2011. Tactile–auditory shape learning engages the lateral occipital complex. J Neurosci. 31:7848–7856.

Lacey S, Flueckiger P, Stilla R, Lava M, Sathian K. 2010. Object familiarity modulates the relationship between visual object imagery and haptic shape perception. Neuroimage. 49:1977–1990.

Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT. 2004. Semantic congruence is a critical factor in multisensory behavioral performance. Exp Brain Res. 158:405–414.
Martin A. 2007. The representation of object concepts in the brain. Annu Rev Psychol. 58:25–45.
Martin A, Chao LL. 2001. Semantic memory and the brain: structure and processes. Curr Opin Neurobiol. 11:194–201.
Meredith MA. 2002. On the neuronal basis for multisensory convergence: a brief overview. Brain Res Cogn Brain Res. 14:31–40.
Mesulam MM. 1998. From sensation to cognition. Brain. 121:1013–1052.
Meyer K, Damasio A. 2009. Convergence and divergence in a neural architecture for recognition and memory. Trends Neurosci. 32:376–382.
Naumer MJ, Doehrmann O, Muller NG, Muckli L, Kaiser J, Hein G. 2009. Cortical plasticity of audio-visual object representations. Cereb Cortex. 19:1641–1653.
Naumer MJ, Ratz L, Yalachkov Y, Polony A, Doehrmann O, van de Ven V, Muller NG, Kaiser J, Hein G. 2010. Visuohaptic convergence in a corticocerebellar network. Eur J Neurosci. 31:1730–1736.
Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ. 2008. The effect of prior visual information on recognition of speech and sounds. Cereb Cortex. 18:598–609.
Noppeney U, Price CJ. 2003. Functional imaging of the semantic system: retrieval of sensory-experienced and verbally learned knowledge. Brain Lang. 84:120–133.
Noppeney U, Price CJ. 2004. Retrieval of abstract semantics. Neuroimage. 22:164–170.
O'Craven KM, Kanwisher N. 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. J Cogn Neurosci. 12:1013–1023.
Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9:97–113.
Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD. 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage. 10:15–35.
Price CJ. 2000. The anatomy of language: contributions from functional neuroimaging. J Anat. 197:335–359.
Reed CL, Klatzky RL, Halgren E. 2005. What vs. where in touch: an fMRI study. Neuroimage. 25:718–726.
Renier LA, Anurova I, De Volder AG, Carlson S, VanMeter J, Rauschecker JP. 2009. Multisensory integration of sounds and vibrotactile stimuli in processing streams for "what" and "where". J Neurosci. 29:10950–10960.
Schneider TR, Lorenz S, Senkowski D, Engel AK. 2011. Gamma-band activity as a signature for cross-modal priming of auditory object recognition by active haptic exploration. J Neurosci. 31: 2502–2510.
Schurmann M, Caetano G, Hlushchuk Y, Jousmaki V, Hari R. 2006. Touch activates human auditory cortex. Neuroimage. 30:1325–1331.
Sestieri C, Di Matteo R, Ferretti A, Del Gratta C, Caulo M, Tartaro A, Olivetti Belardinelli M, Romani GL. 2006. "What" versus "where" in the audio-visual domain: an fMRI study. Neuroimage. 33:672–680.
Stein BE, Stanford TR. 2008. Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci. 9:255–266.
Stevenson RA, James TW. 2009. Audio-visual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. Neuroimage. 44:1210–1223.
Stevenson RA, Kim S, James TW. 2009. An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. Exp Brain Res. 198:183–194.
Tal N, Amedi A. 2009. Multisensory visual-tactile object related network in humans: insights gained using a novel crossmodal adaptation approach. Exp Brain Res. 198:165–182.
Tanabe HC, Honda M, Sadato N. 2005. Functionally segregated neural substrates for arbitrary audio-visual paired-association learning. J Neurosci. 25:6409–6418.
van Atteveldt NM, Blau VC, Blomert L, Goebel R. 2010. fMR-adaptation indicates selectivity to audio-visual content congruency in distributed clusters in human superior temporal cortex. BMC Neurosci. 11:11.
Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL. 1998. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science. 281:1188–1191.
Wheatley T, Weisberg J, Beauchamp MS, Martin A. 2005. Automatic priming of semantically related words reduces activity in the fusiform gyrus. J Cogn Neurosci. 17:1871–1885.
Wheeler ME, Petersen SE, Buckner RL. 2000. Memory's echo: vivid remembering reactivates sensory-specific cortex. Proc Natl Acad Sci U S A. 97:11125–11129.