ELSEVIER

Neuropsychologia 45 (2007) 531–539

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Retinotopic effects during spatial audio-visual integration

A. Meienbrock[a,1], M.J. Naumer[a,b,c,*,1], O. Doehrmann[c], W. Singer[a,b], L. Muckli[a,b]

[a] Max Planck Institute for Brain Research, Department of Neurophysiology, Frankfurt/Main, Germany
[b] Brain Imaging Center (BIC), Frankfurt/Main, Germany
[c] Institute of Medical Psychology, Frankfurt Medical School, Frankfurt/Main, Germany

Received 5 September 2005; received in revised form 5 May 2006; accepted 16 May 2006
Available online 23 June 2006

Abstract

The successful integration of visual and auditory stimuli requires information about whether visual and auditory signals originate from corresponding places in the external world. Here we report crossmodal effects of spatially congruent and incongruent audio-visual (AV) stimulation. Visual and auditory stimuli were presented from one of four horizontal locations in external space. Seven healthy human subjects had to assess the spatial fit of a visual stimulus (i.e. a gray-scaled picture of a cartoon dog) and a simultaneously presented auditory stimulus (i.e. a barking sound). Functional magnetic resonance imaging (fMRI) revealed two distinct networks of cortical regions that processed preferentially either spatially congruent or spatially incongruent AV stimuli. Whereas earlier visual areas responded preferentially to incongruent AV stimulation, higher visual areas of the temporal and parietal cortex (left inferior temporal gyrus [ITG], right posterior superior temporal gyrus/sulcus [pSTG/STS], left intraparietal sulcus [IPS]) and frontal regions (left pre-central gyrus [PreCG], left dorsolateral pre-frontal cortex [DLPFC]) responded preferentially to congruent AV stimulation. A position-resolved analysis revealed three robust cortical representations for each of the four visual stimulus locations in retinotopic visual regions corresponding to the representation of the horizontal meridian in area V1 and at the dorsal and ventral borders between areas V2 and V3. While these regions of interest (ROIs) did not show any significant effect of spatial congruency, we found subregions within ROIs in the right hemisphere that showed an incongruency effect (i.e. an increased fMRI signal during spatially incongruent compared to congruent AV stimulation). We interpret this finding as a correlate of spatially distributed recurrent feedback during mismatch processing: whenever a spatial mismatch is detected in multisensory regions (such as the IPS), processing resources are re-directed to low-level visual areas.

Keywords: Crossmodal; Multisensory; Audio-visual; Functional magnetic resonance imaging (fMRI); Spatial congruency; Retinotopy

1. Introduction

Object perception can involve information from different modalities, thereby engaging several streams of processing. Although these streams operate largely independently at least at early processing stages, visual, auditory, and tactile cues are eventually integrated to form a unified percept of the object. From the large set of possible parameters which might facilitate this kind of multisensory integration, research has focused on temporal synchronicity, as well as spatial and semantic congruency as the crucial dimensions (see e.g. Calvert & Thesen, 2004). Unlike 'natural' object perception which is characterized by a high coincidence of these factors, experimental setups serve to investigate the specific impact of particular stimulus dimensions. Here, we studied the variation of spatial audio-visual (AV) congruency.

The importance of "temporal" and "spatial" congruency has been investigated in animal studies using invasive electrophysiology (for an early review see Stein & Meredith, 1993). These pioneering experiments revealed several features of particular neurons – so-called multisensory integrative (MSI) cells – whose activity is assumed to play a key role for multisensory integration. The most striking among their features is the super-additive response property, i.e. a response to (coincident and/or congruent) bimodal stimulation that exceeds the linear sum of the neuron's responses to the respective unimodal stimuli. MSI cells have been found in a number of species and in both cortical (Wallace, Ramachandran, & Stein, 2004) and sub-cortical brain structures as early as the superior colliculus (Meredith & Stein, 1996).

[*] Corresponding author at: Institut für Medizinische Psychologie, Klinikum der Johann Wolfgang Goethe-Universität, Heinrich-Hoffmann-Str. 10, D-60528 Frankfurt, Germany. Tel.: +49 69 6301 6581; fax: +49 69 6301 7606.
E-mail address: M.J.Naumer@med.uni-frankfurt.de (M.J. Naumer).
[1] These authors contributed equally to this work.

0028-3932/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2006.05.018

In humans, spatial AV integration has been investigated predominantly with psychophysical (e.g. McDonald, Teder-Sälejärvi, & Hillyard, 2000) and electrophysiological methods such as electro- or magnetoencephalography (EEG or MEG, for a review see Eimer & Driver, 2001). Recent ERP findings (Teder-Sälejärvi, Di Russo, McDonald, & Hillyard, 2005) point towards different cortical activation patterns for spatially congruent versus incongruent AV stimuli. However, the spatial resolution of EEG/MEG has been insufficient to identify the cortical sites involved in the processing of these stimuli. Here functional magnetic resonance imaging (fMRI) was used whose high spatial resolution serves to overcome this limitation.

Several imaging studies have identified different general networks of cortical regions that are involved in semantic (Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005) and spatial crossmodal integration (Macaluso & Driver, 2005). Until now, most imaging studies of spatial crossmodal integration have employed visuo-tactile stimuli, while similar investigations in the AV domain are relatively rare. In a recent positron emission tomography (PET) study, Macaluso, George, Dolan, Spence, and Driver (2004) varied both the spatial congruency and temporal synchronicity of auditory and visual stimuli (spoken words and faces pronouncing the same words). Spatially incongruent blocks were associated with activity in more dorsal occipital areas, while ventral occipital areas and the superior temporal sulcus (STS) were unaffected by relative location. In addition, both lateral and dorsal occipital areas were selectively activated by synchronous bimodal stimulation at the same spatial location. The spatial resolution of PET might have been insufficient to detect differential effects in retinotopically organized visual areas. The present study investigated spatial congruency and incongruency effects on retinotopic visual areas by presenting auditory and visual stimuli either at identical or different locations in external space.

2. Methods

2.1. Subjects

Seven subjects (four females, one left-hander) participated in the fMRI experiment; their mean age was 26.0 years (range 23–33 years). All subjects had normal or corrected-to-normal vision (four subjects). All participants received information on MRI and a questionnaire to check for potential health risks and contraindications. Volunteers gave their informed consent after having been introduced to the procedure in accordance with the declaration of Helsinki.

2.2. Stimuli

The visual stimulus was generated by a notebook (DELL 7500, 750 MHz) at a frame rate of 60 Hz. The image was projected onto a vertical screen positioned inside the scanner with an LCD projector (Sony, VPL PX 20) equipped with a custom-made lens. Subjects viewed the screen through a mirror. Mirror and projection screen were fixed onto the head coil. The subjects' field of view was 50° visual angle (maximum horizontal distance).

The visual stimulation (Fig. 1) consisted of four gray-scaled pictures of cartoon dogs, one of which was located peripherally and one centrally in each visual hemifield. They were adjusted in size to compensate for the cortical magnification factor (Sereno et al., 1995). All dogs were positioned along the horizontal meridian. Information on sizes and positions of the visual stimuli is provided in Table 1. In the center of the screen, a fixation cross was presented during the whole experiment. The screen color was blue. Employing a classical block design, only one dog was presented during a stimulation block. The black-and-white dogs inverted their contrast every 100 ms (checkerdog) to obtain maximal visual stimulation. Eight experimental conditions (lasting 12 s

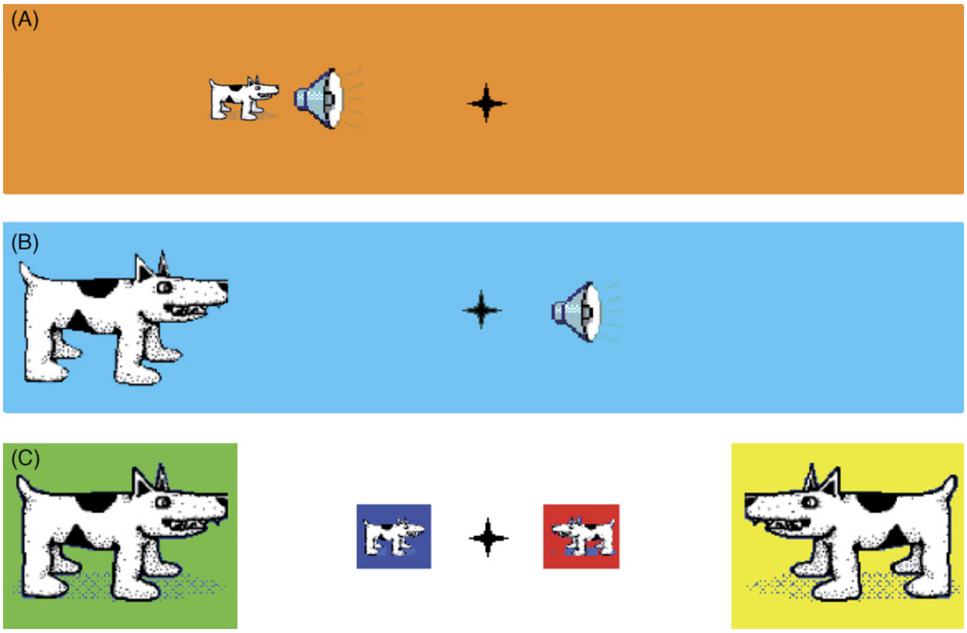

Fig. 1. Experimental design. Audio-visual (AV) stimulation consisted of gray-scaled pictures of a cartoon dog and barking sounds. Auditory and visual stimuli were presented synchronously, but could be either spatially congruent (A) or incongruent (B), depending on whether the unimodal stimulus components were presented at the same or at two different locations. In the incongruent conditions, the auditory stimulus was displaced from the visual stimulus by 20° visual angle towards the contralateral hemi-space. (C) Illustrates the four locations in external space where stimuli could occur. We employed a block design with eight experimental conditions (i.e. 2 degrees of spatial congruency for each of four locations). Size of laterally and medially presented cartoon dogs differed to compensate for the cortical magnification factor (Sereno et al., 1995). Information on sizes and positions of the visual stimuli is provided in Table 1.

Table 1
Sizes and positions of the visual stimuli

	L	LM	RM	R
Length	11°	3°	3°	11°
Height	7.3°	2°	2°	7.3°
Horizontal position	−16.5°	−3.5°	+3.5°	+16.5°

L, left; LM, left medial; RM, right medial; R, right locations.

per stimulation block) were repeated four times each during the experiment in a pseudo-randomized order, alternating with fixation periods (20 s). The entire experiment thus consisted of 32 stimulation periods alternating with 33 fixation periods, and lasted 17 min and 20 s.

Auditory stimuli were generated by the same notebook and passed through an amplifier (JVC AX R5BK). Sounds were presented via MR-compatible headphones which were fitted into an earmuff (Bilsom 727, sound insulation of 33 dB at 1000 Hz). We built these headphones by removing the permanent magnets of standard stereo headphones (impedance of 32 Ω, frequency ranging from 20 till 20,000 Hz) and using the external magnetic field of the MR tomograph. In general, sound quality was superior to conventional MR-compatible stimulation systems based on air-conducting tubes.

The auditory stimulation consisted of four spatially different barking sounds, mapped on the four visual stimulus locations by varying interaural intensity differences. The spatial fit to the four locations was assessed in a separate behavioral experiment prior to the fMRI experiment. Here, three of our subjects judged 20 different sounds with respect to their spatial fit to the four visual stimulus locations. These 20 sounds were created by using a wave-file of a barking dog (with equal volume for both channels) and adjusting the volume of the right and left channels systematically in steps of 5% in opposite directions.

This resulted in intracranial percepts with deviations from the midsaggital plane that ranged between +90° and −90° (Zimmer, Lewald, Erb, Grodd, & Karnath, 2004). Interaural time difference remained unchanged. The four sound positions rated closest to each of the visual locations were used for the fMRI experiment. Stimulation during each of our experimental conditions consisted of the respective barking sound (length ∼1s) that was repeated 12 times per block. During the experiment, each of the four auditory stimuli was presented during eight blocks of the same experimental run. In half of these cases the resulting AV stimulation was spatially congruent (Fig. 1A). During spatially incongruent stimulation (Fig. 1B), the distance between the respective auditory and visual stimulus locations remained constant and amounted to 20° of visual angle. The auditory stimuli were displaced towards the respective contralateral hemi-space. Auditory and visual stimulation onsets were synchronous.

2.3. Procedure

Subjects were instructed to pay attention to the experimental stimuli while maintaining central fixation in order to minimize eye-movements. The subjects' task was to indicate via button-press, whether AV stimulation was perceived to be spatially congruent or incongruent.

2.4. Behavioral training

In a separate training session outside the tomograph, subjects were seated in a chair wearing conventional stereo headphones (Philips SBC HS 400/00), with their head position restrained by a chin rest. Distance between the subject's eyes and the screen of the notebook (DELL 7500, 750 MHz) was 33 cm, resulting in the same visual angle of visual stimulation as inside the tomograph. Subjects had to decide about the spatial congruence of the AV stimulation and were trained until individual performance was well above chance level.

2.5. Control of eye movements

To control for confounding stimulus-related eye movements, we used an infrared eye tracker to monitor eye movements in three of our subjects during the behavioral training session outside the tomograph. We found no detectable pattern of eye movements correlated with the locations of visual stimulation.

2.6. Imaging

FMRI data acquisition was performed using a 1.5 T Siemens Magnetom Vision tomograph (Siemens, Erlangen, Germany) at the Institute of Neuroradiology, University Hospital, Johann Wolfgang Goethe-University, Frankfurt am Main. A gradient-recalled echo-planar-imaging (EPI) sequence was used with the following parameters: 16 slices, oriented approximately in parallel to the AC–PC plane (AC, anterior commisure; PC, posterior commissure); TR, 2081 ms; TE, 69 ms; FA, 90°; FOV, 200 mm; in-plane resolution, 3.13 mm × 3.13 mm; slice thickness, 5 mm; gap thickness, 1 mm. In addition, a detailed T1-weighted anatomical scan was acquired for all subjects using a Siemens fast low-angle-shot (FLASH) sequence (isotropic voxel size 1 mm^3). For each subject, a magnetization-prepared rapid-acquisition gradient-echo (MP-RAGE) sequence was used (TR = 9.7 ms, TE = 4 ms, FA = 12°, matrix = 256 × 256, voxel size 2.0 mm × 1.0 mm × 1.0 mm) for realignment with the detailed anatomical images that had been acquired in a previous session.

2.7. Data analysis

Data were analyzed using the BrainVoyager^{TM} 2000 (version 4.9) and BrainVoyager^{TM} QX (version 1.6) software packages (Rainer Goebel, Brain Innovation, Maastricht, The Netherlands, http://www.brainvoyager.com). The first four volumes of each scan were discarded to preclude T1 saturation effects. Preprocessing of the functional data included the following steps: (i) three-dimensional motion correction, (ii) linear-trend removal and temporal high-pass filtering at 0.01 Hz, (iii) slice-scan-time correction with sinc interpolation, and (iv) the alignment of individual cortices to one another using an algorithm accounting for an optimal fit of the main gyrification with minimal distortion between the individual cortices (see below). Cortex-based statistical analysis was performed using multiple linear regression. For every cortical surface vertex, the time course was regressed on a set of dummy-coded predictors representing our eight experimental conditions. To account for the shape and delay of the hemodynamic response (Boynton, Engel, Glover, & Heeger, 1996), the predictor time courses (box-car functions) were convolved with a gamma function. A contrast-based statistical analysis was performed, using a *t*-test. We thereby defined sets of predictors (experimental conditions) for each of our subjects individually (so-called "separate subject predictors"), and employed all of them during multiple linear regression to obtain our group-statistics.

In a first step, we revealed the general cortical network involved in spatial AV integration. Then, applying a position-resolved analysis, we further searched for modulatory effects in retinotopically organized low-level visual cortices. During the initial location-spanning analysis, we compared all four spatially congruent to all four incongruent AV conditions (resulting contrasts: *congruent > incongruent* and *incongruent > congruent*, respectively). Vertices were included into the statistical maps if the obtained *p*-value was <0.01 (uncorrected). We performed separate analyses based on either all trials ("physically" congruent versus incongruent) or correct trials only ("corrected" analyses). In a second step, we performed a balanced contrast of one spatial location with the other three locations. Vertices were included into the statistical map if the obtained *p*-value was <0.0001 (corrected). Finally, we contrasted the spatially congruent and incongruent conditions separately for each of the four locations. Vertices were included into the respective statistical map if the obtained *p*-value was <0.001 (uncorrected).

During analysis, statistical maps of group data were superimposed on inflated hemispheres of one subject (DL). The response profiles for each of our regions of interest (ROIs) were visualized by plotting either the event-related BOLD-signal averages or the GLM beta weights for the different experimental conditions.

2.8. Retinotopic mapping

Phase-encoded retinotopic mapping was assessed in each subject and included mapping of eccentricity and polar angle (Engel et al., 1994; Goebel, Khorram-Sefat, Muckli, Hacker, & Singer, 1998; Muckli, Kohler, Kriegeskorte, & Singer, 2005; Sereno et al., 1995). In the eccentricity mapping experiment, black and white checkerboard patterns were presented in a ring-shaped configuration and were flickered at a rate of 4 Hz. The ring started with a radius of 1° and slowly expanded to a radius of 12° visual angle within 64 s. In the polar-angle mapping experiment, the checkerboard pattern consisted of a ray-shaped disk segment subtending 22.5° of polar angle. The ray started at the right horizontal meridian and slowly rotated clockwise for a full cycle of 360° (within 64 s). Each mapping experiment consisted of seven repetitions of a full expansion (each cycle lasting for 64 s) or 10 repetitions of rotation, respectively.

The analysis of the retinotopic-mapping experiment was conducted by the use of a cross-correlation analysis. We used the predicted hemodynamic signal time course for the first 1/8 of a stimulation cycle (corresponding to 45° visual angle in the polar mapping experiment) and shifted this reference function successively in time (time steps corresponded to the recording time for one volume, TR; see also Muckli et al., 2005). Sites activated at particular eccentricities and polar angles were identified through selection of the lag value that resulted in the highest cross-correlation value for a particular voxel. The obtained lag values at particular voxels were encoded in pseudo-color on corresponding surface patches (triangles) of the reconstructed cortical sheet. Based on the polar-angle mapping experiment, the boundaries of retinotopic cortical areas V1, V2, V3, V3A, and V4v were estimated manually on the inflated cortical surface.

2.9. Cortical-surface reconstruction and visualization

The high-resolution T1-weighted 3D recordings were used for surface reconstruction of both cerebral hemispheres of each subject (Kriegeskorte & Goebel, 2001). The white/gray-matter border was segmented with a region-growing method preceded by inhomogeneity correction of signal intensity across space. The borders of the two resulting segmented subvolumes were tessellated to produce a surface reconstruction of each hemisphere.

2.10. Cortex-based inter-subject alignment

To improve the spatial correspondence between subjects' brains beyond Talairach space matching, the reconstructed hemispheres were aligned using curvature information reflecting the gyral/sulcal folding pattern (see van Atteveldt, Formisano, Goebel, & Blomert, 2004 for details). As demonstrated recently for a similar analysis tool (Argall, Saad, & Beauchamp, 2006), such a surface-based approach can help to improve *t*-statistics in the average activation map as compared to the volume average.

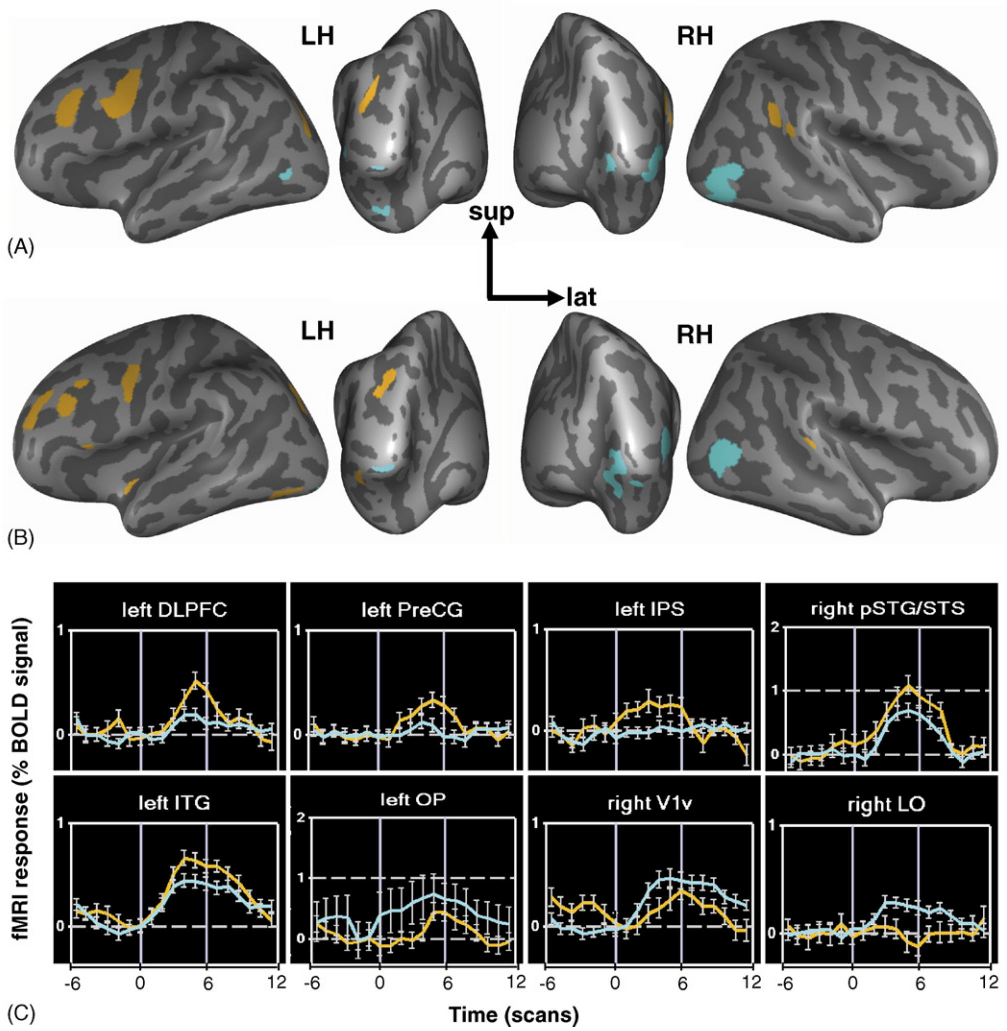

Fig. 2. Cortical regions sensitive to spatial congruency. (A) Analysis based on physical stimulation. Inflated representations are depicted of one subject's cerebral hemispheres in lateral (lateral panels) and posterior views (medial panels). Statistical maps for the contrasts (*physically*) *congruent* > *incongruent* (in orange) and (*physically*) *incongruent* > *congruent* AV stimulation (in light blue) were based on multiple regression of group-averaged data ($n = 7$). (B) Analysis based on correct trials only. Inflated representations are depicted of one subject's cerebral hemispheres in lateral (lateral panels) and posterior views (medial panels). Statistical maps for the contrasts *congruent* > *incongruent* (in orange) and *incongruent* > *congruent* AV stimulation (in light blue) were based on multiple regression of group-averaged data ($n = 6$). (C) Shows the mean BOLD-signal intensity changes during perception of spatially congruent (in orange) and incongruent AV stimulation (in light blue) for eight selected cortical regions from (B). All maps were significant at $p < 0.01$ (uncorrected).

Folded cortical representations of each subject and hemisphere were morphed into a spherical representation that provided a parameterizable surface for nonrigid alignment across subjects. The curvature information of the folded representation was preserved as a curvature map on the spherical representation. This folding pattern was smoothed along the sphere surface to provide spatially extended gradient information driving inter-subject alignment. Following a coarse-to-fine matching strategy, the alignment started with highly smoothed curvature maps and progressed to only slightly smoothed representations. While the alignment of major gyri and sulci was achieved reliably using this method, smaller structures, reflecting idiosyncratic differences between the subjects' brains, were not completely aligned.

Cortex-based inter-subject alignment enabled us to align the time courses for multisubject GLM data analysis. Group-averaged functional data were then projected on inflated representations of the left and right cerebral hemispheres of a single subject (DL; see Fig. 2A). As a morphed surface always possesses a link to the folded reference mesh, functional data can be shown at the correct location of folded as well as inflated representations. This link was also used to keep geometric distortions to a minimum during inflation through inclusion of a morphing force that keeps the distances between vertices and the area of each triangle of the morphed surface as close as possible to the respective values of the folded reference mesh.

3. Results

At the behavioral level, subjects had classified on average 85.5% of all trials correctly, but incongruent trials (92.2% correct) were detected with a substantially higher accuracy than congruent trials (57.5% correct). As subjects' performance was found to be markedly worse in the congruent condition, we conducted sensitivity analyses (d' analyses) for each of our subjects. Table 2 lists these individual d'-scores that ranged between 0.68 and 3.13 (with a mean d' [$n = 6$] of 1.83). Three subjects (BV, GS, and MN) performed rather poorly and reached only low sensitivity values ($d' < 1.3$). Moreover it seems that these subjects were employing a conservative response criterion that prevented a positive (i.e. 'congruent') response except in cases of high certainty. In order to use perceptually pure conditions we only used certified trials in later steps of data analysis (corrected).

Presentation of AV stimuli activated a distributed network of cortical regions in the occipital, temporal, parietal, and frontal lobes (Fig. 2). In a location-spanning analysis, we could reveal two independent networks whose activation reflected the degree of spatial AV congruency and incongruency in a complementary manner. Statistical maps for the contrasts *congruent > incongruent* and *incongruent > congruent* AV stimulation were based on multiple regression of group-averaged data. Separate analyses were based either on all trials (seven subjects; Fig. 2A) or on correct trials only (six subjects; Fig. 2B) and resulted in largely similar cortical activation maps ($p < 0.01$, uncorrected). Regions showing a preference for spatial AV congruency (see orange-colored regions in Fig. 2) included left inferior temporal gyrus (ITG), intra-parietal sulcus (IPS), pre-central gyrus (PreCG), and dorsolateral pre-frontal cortex (DLPFC), and right posterior superior temporal gyrus/sulcus (pSTG/STS). On the other hand, retinotopic low-level visual regions (including V1) in bilateral occipital cortex showed a significantly higher BOLD-signal during spatially incongruent as compared to congruent AV stimulation (see light blue-colored regions in Fig. 2).

Table 2
Behavioral sensitivity data (individual d' analyses)

Subject	d'
AF	2.45
BV	1.05
CM	3.13
DL	2.45
GS	1.22
MN	0.68
$n = 6$	Mean = 1.83

In a second step, our goal was to find separable foci of BOLD activation that corresponded to the four different locations along the horizontal meridian of the visual field. In order to carry out an analysis of single-subject data, we performed a retinotopic mapping procedure (see Section 2 for details) and defined the borders of retinotopic low-level visual areas for each subject individually. The horizontal meridian (i.e. where we presented our experimental stimuli) is represented at three distinct locations within human retinotopic visual areas V1–V3: in the middle of V1, at the border between dorsal V2 and V3, and between ventral V2 and V3. Using a surface-based analysis of the individual data sets, we identified those regions showing a BOLD-signal increase in response to our visual stimuli and overlapping with the respective cortical representations of the horizontal midline. We were able to separate foveal from peripheral activation by contrasting the response to each of our four visual stimulus locations [left (L), left-middle (LM), right-middle (RM), right (R)] and pooling the respective congruent and incongruent conditions for each of the locations. The mapping results are shown based on group-averaged data ($n = 7$) in Fig. 3. For each of our visual stimulus locations, cortex-based group analysis revealed three distinct representations in the contralateral hemisphere. One of these representations consisted of surface patches extending from the border between dorsal V2 and V3 to the border between ventral V2 and V3, including parts of V1. The two other representations were located more laterally in LO (tentatively labeled *lateral occipital*) and dorsally around the IPS (see lower panels in Fig. 3 for the respective response profiles). While all these ROIs showed a highly significant location preference ($p < 0.0001$, corrected), none of them showed a significant preference for spatially congruent or incongruent AV stimulation at the respective location of the visual field.

In a final step, we directly contrasted the spatially congruent and incongruent experimental conditions separately for each of our four visual stimulus locations. Only in the right hemisphere were we able to reveal subdivisions within the above reported visual location ROIs that showed a significant BOLD-signal increase in response to spatially incongruent as compared to congruent AV stimulation ($p < 0.001$, uncorrected; Fig. 4).

4. Discussion

Employing BOLD-fMRI, we were able to reveal two distinct networks of cortical regions preferentially processing either spatially congruent (left ITG, IPS, PreCG, DLPFC, and right pSTG/STS) or spatially incongruent AV stimulation (bilateral retinotopic low-level visual areas). A position-resolved anal-

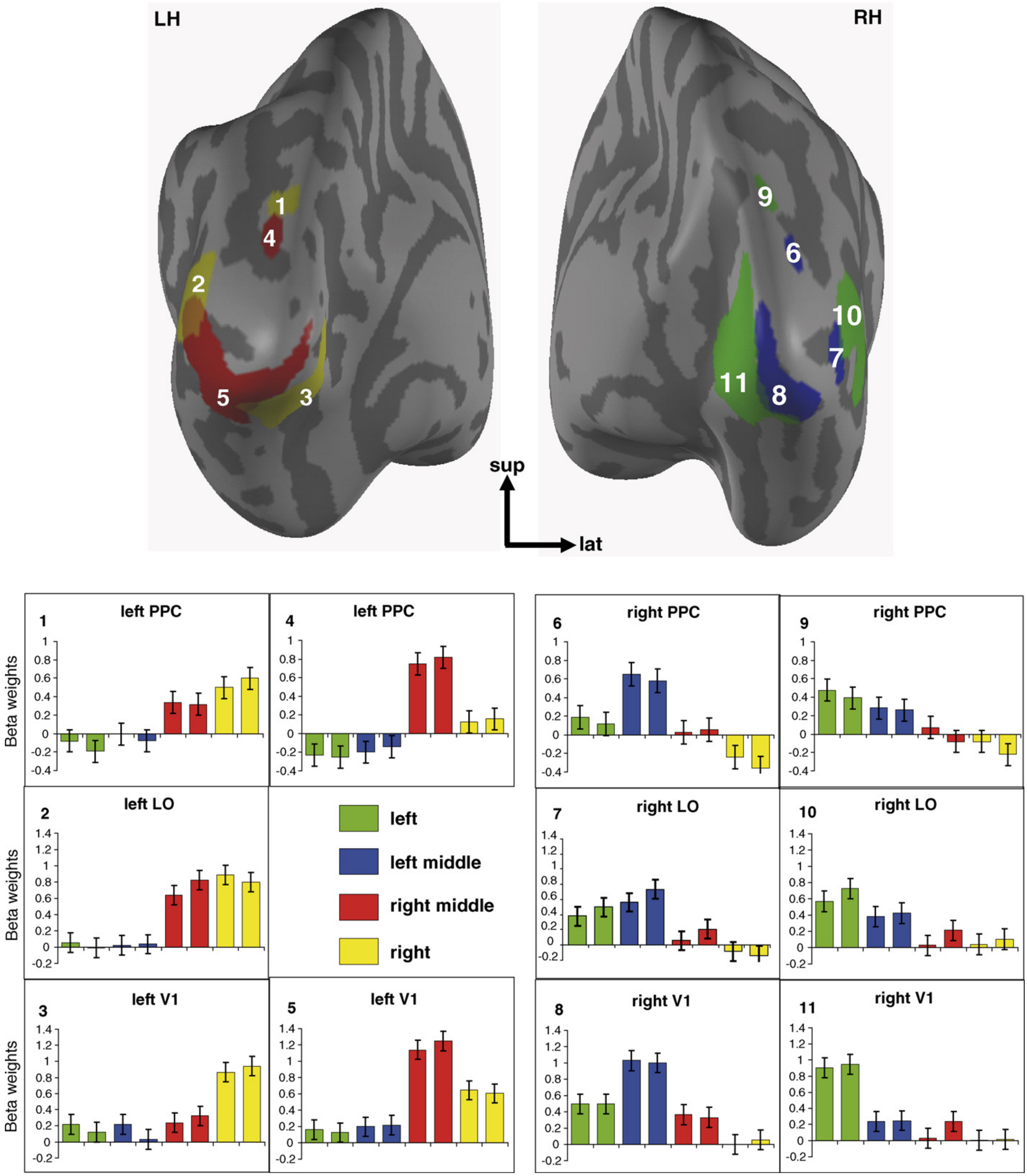

Fig. 3. Mean cortical representations of the four visual stimulus locations. Group-averaged ($n = 7$) location maps were projected on inflated representations of one subject's cerebral hemispheres (DL; shown in upper panels; view from posterior). Colors code for the different locations as follows: yellow, right (R); red, right medial (RM); blue, left medial (LM); green, left (L). For each of these locations, analysis revealed three distinct representations in the contralateral hemisphere. One of these representations consisted of surface patches extending from the border between dorsal V2 and V3 to the border between ventral V2 and V3, including parts of V1. The two other representations were located more laterally in LO (tentative label for *lateral occipital*) and dorsally around the posterior intra-parietal sulcus (IPS). For these ROIs, the mean BOLD-signal intensity changes (beta weights) are shown in response to each of our eight experimental conditions (lower panels). While all ROIs showed a highly significant location preference ($p < 0.0001$, corrected), none of them showed a significant preference for spatially congruent or incongruent AV stimulation.

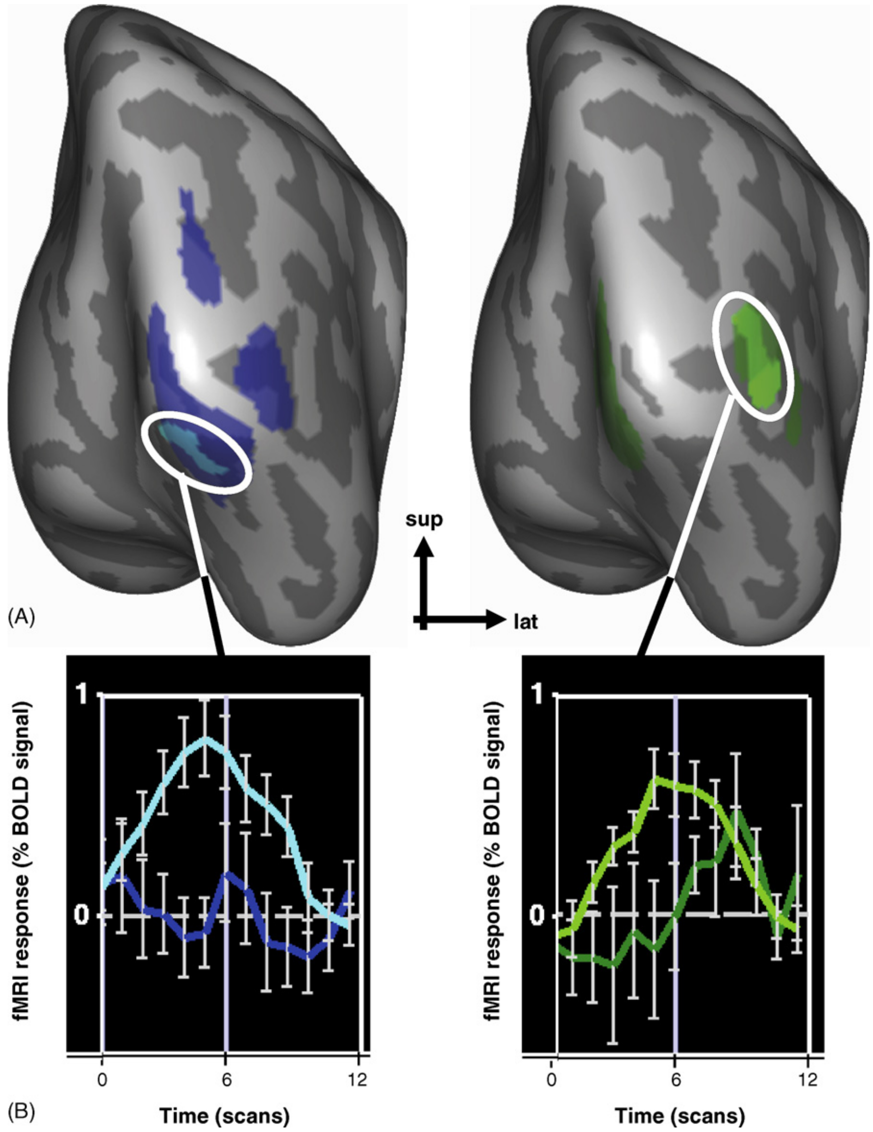

Fig. 4. Retinotopic effects of spatial AV incongruency (right hemisphere only). (A) Group-averaged ($n = 7$) location maps ($p < 0.05$, corrected) were projected onto inflated representations of one subject's right cerebral hemisphere (as viewed from posterior). Colors code for the two different locations within the contralateral visual hemi-field as follows: blue, left medial (LM); green, left (L). In addition, group-averaged ($n = 6$) incongruency maps (as shown in lighter shades of the same colors; contrast: spatially *incongruent > congruent*; $p < 0.001$, uncorrected) have been computed for each of the two locations, separately. These incongruency maps (based on analyses of correct trials) partially overlapped with the respective location maps. (B) For each of the overlapping regions from (A) (circles) mean BOLD-signal intensity changes are shown. Color code for the graphs—light blue: LM incongruent; dark blue: LM congruent; light green: L incongruent; dark green: L congruent.

ysis revealed robust cortical representations for each of our four visual stimulus positions in contralateral visual areas V1, V2d/V3d, V2v/V3v, LO, and around posterior IPS. While these ROIs did not show any significant effect of spatial congruency, we found subregions within right hemisphere ROIs that showed a spatial AV incongruency effect (i.e. a higher BOLD-signal during spatially incongruent as compared to congruent AV stimulation).

4.1. *Effects of spatial AV congruency*

Our findings concerning spatial congruency involve two networks of cortical regions. At a lower level, we found activation in left IPS and right pSTG/STS, which is consistent with a number of other neuroimaging findings (see Amedi et al., 2005; Beauchamp, 2005; Macaluso & Driver, 2005 for recent reviews). As these areas are considered to be important heteromodal (or multisensory) integration sites, they are likely to detect spatial congruency between semantically related auditory and visual stimuli.

The second network, operating at a higher processing level, is constituted by frontal regions. At first glance, activation here could be due to a higher effort of the subjects during detection of *spatial congruency*, which might have resulted in a higher amount of task-directed attention resources (Corbetta & Shulman, 2002). This would be in line with our subjects' behavioral performance differing significantly between the spatially congruent and incongruent conditions.

4.2. Spatial AV incongruency effects

A main result of our study was the observed effect of spatially incongruent AV stimuli in low-level visual regions. Although these regions might not be involved in detecting the spatial mismatch of AV stimulation, they might be affected as a consequence of this detected mismatch. Higher responses to spatially incongruent AV stimuli cannot be explained by intensity differences, since the compared experimental conditions consisted of nearly identical visual and auditory stimuli. The auditory stimuli were counterbalanced with respect to the different visual locations. While the activation of the aforementioned network of frontal regions (during spatially congruent AV stimulation) might indicate an employment of task-directed attention resources (Corbetta & Shulman, 2002), low-level visual areas show an AV incongruency effect. If it were the case that both networks participate in the same task-related operations, we should have found rather a congruency effect in low-level visual areas—which we have not.

In addition, our analyses revealed that subjects exhibited neural response patterns in low-level visual cortices that differed between the left and right hemispheres. While we found a rather robust retinotopic AV incongruency effect in the right hemisphere (as shown in Fig. 4), we were not able to reveal a similar incongruency pattern in the left hemisphere (at least when applying the same statistical thresholds). We assume that this mainly reflects the fact that subjects made substantially more errors during physically congruent AV stimulation within the left (58.3% incorrect) as opposed to the right hemi-space (29.2% incorrect). In other words: subjects had a stronger tendency to perceive our AV stimuli as being spatially incongruent during stimulation within the left hemi-space.

In sum, we conclude that the activation of retinotopic low-level visual areas can neither be explained by stimulus energy differences nor by the specific behavioral task requirements. Thus, the variation of spatial AV congruency is the most likely factor to cause this finding.

Which mechanism might be involved in the processing of spatially incongruent multisensory stimuli? We favor the hypothesis that the activation of low-level visual areas reflects rather automatic and rapid processes to resolve a spatial mismatch that has been detected at the level of heteromodal (or multisensory) regions (such as IPS and pSTG/STS). It is likely that these processes are controlled via feedback (Mesulam, 1998) from the aforementioned two networks—directly by the heteromodal regions and rather indirectly by the supramodal frontal network. This interpretation is supported by findings of studies investigating the timing of crossmodal spatial integration. McDonald, Teder-Sälejärvi, Di Russo, and Hillyard (2003) found a modulation of visual processing by a spatially varying auditory cue after 120–140 ms in the superior temporal cortex and 15–25 ms later in ventral occipital cortex. The authors conclude that this temporal pattern might reflect an attentional modulation of visual regions by recurrent feedback originating in heteromodal lateral temporal regions. Using MEG, Kaiser, Hertrich, Ackermann, Mathiak, and Lutzenberger (2005) reported a top-down driven modulation of low-level visual networks by higher posterior parietal regions during processing of incongruent AV speech. Our interpretation is fully in line with those two studies. What is more, spatial AV incongruency could be interpreted as a salient stimulus, which automatically attracts attention resources to the visual domain (Corbetta & Shulman, 2002; Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999). To resolve the detected incongruency with high speed, the involvement of low- to mid-level stages of processing appears to be advantageous. This level could be constituted by the reported network involving low-level visual areas, left IPS and right pSTG/STS.

Still, it remains unclear why the visual cortex seems to play such a prominent role in the processing of spatial incongruency. With regard to modality dominance, research has generally found that vision is dominant over audition for spatial perception (*information reliability hypothesis*; Schwartz, Robert-Ribes, & Escudier, 1998). In addition, *modality appropriateness* to the respective task (Warren, 1979; Welch & Warren, 1980) determines the direction of the crossmodal effect. In our case of spatial AV mismatch, the superior spatial resolution of low-level visual areas offers the most reliable information for successful conflict resolution.

In addition, one has to take into account that the employed biological stimuli are also semantically congruent. That is, the visual and auditory representations pertain to the same animal and are therefore not necessarily generalizable to all types of audio-visual processing (including, for example, synchronous combinations of abstract checkerboard and white noise stimuli). It is indeed well known from the literature on conceptual representations of living (versus non-living) objects that there is a substantial involvement of low-level visual regions in occipital cortex during processing of animals (compared to tools; e.g. Chao, Haxby, & Martin, 1999). However, it appears to be unlikely that the *semantic AV congruency* continuously present in our study contributed to the reported *spatial AV incongruency* effects, although one could easily imagine a substantial saliency increase for *semantically incongruent AV combinations* (e.g. a barking cow) that might strongly involve similar cortical networks. Considering the small number of subjects, the present findings have to be treated as preliminary. Future studies are warranted to test whether our interpretation can be generalized to different types of congruency and different modalities.

Thus, our favored interpretation can be summarized as follows: feed-forward processing involves unisensory auditory and visual cortices and is reduced whenever a spatial mismatch between visual and auditory stimuli has been detected at the level of multisensory regions (such as the IPS and pSTG/STS). In such a case the superior spatial resolution of the visual system is fully utilized via a feedback loop (down to V1) to re-direct attention resources to resolve the detected incongruency. This resulting interaction between low- and mid-level regions can be further modulated by frontal regions (e.g. DLPFC and PreCG), which enable the adaptation to the requirements of a specific task.

Acknowledgements

This research was funded by the German Federal Ministry for Education and Research (BMBF) and the Max Planck Society

(MPS). We thank Lutz Fiesser for helping to develop the auditory stimulation setup. We are grateful to Axel Kohler, Christian Altmann, Jochen Kaiser, and two anonymous reviewers (for helpful comments on earlier versions of this manuscript).

References

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research*, *166*, 559–571.

Argall, B. D., Saad, Z. S., & Beauchamp, M. S. (2006). Simplified intersubject averaging on the cortical surface using SUMA. *Human Brain Mapping*, *27*(1), 14–27.

Beauchamp, M. S. (2005). See me, hear me, touch me: Multisensory integration in lateral occipital–temporal cortex. *Current Opinion in Neurobiology*, *15*(2), 145–153.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, *16*(13), 4207–4221.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal Physiology Paris*, *98*(1–3), 191–205.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215.

Eimer, M., & Driver, J. (2001). Crossmodal links in endogenous and exogenous spatial attention: Evidence from event-related brain potential studies. *Neuroscience and Biobehavioral Reviews*, *25*(6), 497–511.

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., et al. (1994). FMRI of human visual cortex. *Nature*, *369*(6481), 525.

Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., & Singer, W. (1998). The constructive nature of vision: Direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *European Journal of Neuroscience*, *10*(5), 1563–1573.

Kaiser, J., Hertrich, I., Ackermann, H., Mathiak, K., & Lutzenberger, W. (2005). Hearing lips: Gamma-band activity during audiovisual speech perception. *Cerebral Cortex*, *15*(5), 646–653.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761.

Kriegeskorte, N., & Goebel, R. (2001). An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes. *Neuroimage*, *14*(2), 329–346.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: A window onto functional integration in the human brain. *Trends in Neuroscience*, *28*(5), 264–271.

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: A PET study. *Neuroimage*, *21*(2), 725–732.

McDonald, J. J., Teder-Sälejärvi, W. A., Di Russo, F., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. *Journal of Cognitive Neuroscience*, *15*(1), 10–19.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*(6806), 906–908.

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, *75*(5), 1843–1857.

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*(Pt 6), 1013–1052.

Muckli, L., Kohler, A., Kriegeskorte, N., & Singer, W. (2005). Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS Biology*, *3*(8), e265.

Schwartz, J.-L., Robert-Ribes, J., & Escudier, P. (1998). Ten years after summerfield: A taxonomy of models for audio-visual fusion in speech perception. In R. Campbell (Ed.), *Hearing by eye: The psychology of lipreading* (pp. 3–51). Hove, UK: Lawrence Erlbaum Associates.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–893.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Teder-Sälejärvi, W. A., Di Russo, F., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience*, *17*, 1396–1409.

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*(2), 271–282.

Wallace, M. T., Ramachandran, R., & Stein, B. E. (2004). A revised view of sensory cortical parcellation. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(7), 2167–2172.

Warren, D. H. (1979). Spatial localization under conflict conditions: Is there a single explanation? *Perception*, *8*(3), 323–337.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667.

Zimmer, U., Lewald, J., Erb, M., Grodd, W., & Karnath, H. O. (2004). Is there a role of visual cortex in spatial hearing? *European Journal of Neuroscience*, *20*(11), 3148–3156.