Subcortical, Modality-Specific Pathways Contribute to Multisensory Processing in Humans

R. L. van den Brink[1], M. X. Cohen[1,2], E. van der Burg[3], D. Talsma[4], M. E. Vissers[1,2] and H. A. Slagter[1,2*]

[1]Department of Psychology and [2]Cognitive Science Center Amsterdam, University of Amsterdam, Amsterdam, 1081XA, The Netherlands, [3]School of Psychology, University of Sydney, Sydney, NSW 2006, Australia and [4]Department of Experimental Psychology, Ghent University, Ghent, B-9000, Belgium

*Address correspondence to Dr H. A. Slagter, Weesperplein 4, 1081XA Amsterdam, The Netherlands. Email: h.a.slagter@uva.nl

Oftentimes, we perceive our environment by integrating information across multiple senses. Recent studies suggest that such integration occurs at much earlier processing stages than once thought possible, including in thalamic nuclei and putatively unisensory cortical brain regions. Here, we used diffusion tensor imaging (DTI) and an audiovisual integration task to test the hypothesis that anatomical connections between sensory-related subcortical structures and sensory cortical areas govern multisensory processing in humans. Twenty-five subjects (mean age 22 years, 22 females) participated in the study. In line with our hypothesis, we show that estimated strength of white-matter connections between the first relay station in the auditory processing stream (the cochlear nucleus), the auditory thalamus, and primary auditory cortex predicted one's ability to combine auditory and visual information in a visual search task. This finding supports a growing body of work that indicates that subcortical sensory pathways do not only feed forward unisensory information to the cortex, and suggests that anatomical brain connectivity contributes to multisensory processing ability in humans.

Keywords: brain, diffusion tensor imaging, multisensory, multisensory integration, structural connectivity

Introduction

Perception is fundamentally a multisensory experience: We often simultaneously hear and see someone speak; we smell our food as we taste it; and so forth. Nevertheless, research has traditionally studied the senses in isolation, leading to the view that cross-modal input is initially processed in anatomically separate brain regions and pathways (for reviews, see Schroeder and Foxe 2005; Macaluso 2006). In recent years, research in animals (Schroeder and Foxe 2002; Ghazanfar et al. 2005; Ghazanfar and Schroeder 2006; Kayser et al. 2008) as well as humans (Giard and Péronnet 1999; Molholm et al. 2002; Macaluso 2006; Martuzzi et al. 2006; Kayser and Logothetis 2007; Noesselt et al. 2007; Van der Burg et al. 2011) has shown, however, that multisensory interactions do not only occur in multisensory convergence zones in higher-level association cortex (Beauchamp et al. 2004; Ghazanfar and Schroeder 2006; Stein and Stanford 2008), but also at cortical information processing stages as early as the primary sensory cortex. For example, human event-related potential (ERP) studies have reported multisensory interactions as early as 40–60 ms over sensory scalp regions (Giard and Péronnet 1999; Molholm et al. 2002). In addition, human neuroimaging studies have provided some evidence that subcortical structures contribute to multisensory processing as well. For instance, Noesselt et al. (2010) showed that functional connectivity between the medial geniculate nucleus (MGN) of the thalamus and auditory cortex, and between the lateral geniculate nucleus (LGN) and visual cortex, was modulated by multisensory processing and predicted multisensory task performance. Multisensory interactions may also occur in the colliculi and other brainstem structures (Fort et al. 2002; Musacchia et al. 2006; Fairhall and Macaluso 2009).

Invasive histological tracing in animals suggests that multisensory interactions may also be governed by "anatomical" connectivity between modality-specific pathways (for a recent review, see, e.g., Cappe et al. 2009). For example, the inferior colliculus is the primary relay nucleus for auditory input into the MGN, yet it also receives monosynaptic retinal innervations in rats, cats, and monkeys (Itaya and Van Hoesen 1982). The MGN furthermore receives input from the superior colliculus (Benevento and Fallon 1975; Linke 1999), the first relay system within the visual system. Additionally the inferior and superior colliculi are interconnected (Benevento and Fallon 1975). Moreover, there is a direct, short-latency projection from the cochlear nucleus to MGN (Malmierca et al. 2002; Anderson et al. 2006), which parallels the classical lemniscal auditory pathway, and which connects auditory input with inputs from the visuomotor system and other nonauditory systems. Animal studies have further demonstrated the presence of multisensory activity in different subcortical structures (Stein and Stanford 2008), including MGN (Komura et al. 2005). Yet, the functional consequences of anatomical connections between putatively modality-specific pathways are largely unknown. Also, there are large differences in thalamic organization across species (Jones 2007), and it is presently unknown to what extent subcortical anatomical pathways govern multisensory processing in humans. Addressing this question is important, because a fundamental understanding of the extend to which these pathways contribute to multisensory processing ability may have significant ramifications for ongoing research aimed at resolving the complex interactions between various brain areas in humans contributing to multisensory processing (e.g., Driver and Noesselt 2008; Kayser et al. 2008; van Atteveldt et al. 2010).

The current study examined whether anatomical connections between subcortical and cortical sensory brain regions govern multisensory processing in humans. To this end, an audiovisual integration task (to index multisensory processing ability) and diffusion tensor imaging (DTI) were used. DTI enables probabilistic reconstruction of white-matter tracts (or probabilistic tractography) in vivo, based on voxelwise values of anisotropic diffusion (Mori and Zhang 2006). Specifically, probabilistic tractography determines the route of least hindrance to diffusion and permits estimation of interconnection between brain regions. From each subject, we obtained DTI and fMRI auditory/visual localizer data, and, in a separate session, performance data on a visual search task in which nonspatial auditory stimuli (short tone pips) could be used to facilitate detection of

a visual target (a horizontal or vertical line segment) presented among visual distractors (line segments of a different orientation) (see Fig. 1; cf., Van der Burg et al. 2008b). For each individual, we quantified the benefit of the auditory signal on visual search efficiency (or multisensory processing ability), and estimated connection strength between fMRI-defined auditory and visual cortical regions and subcortical structures within, respectively, the auditory and the visual system using probabilistic tractography. We predicted that connection strength within these sensory systems would at least in part determine multisensory processing ability.

Materials and Methods

Participants

Twenty-five healthy participants (22 females; mean age 21.76 years; SD 5.23 years) from the University of Amsterdam participated in the experiment. Participants were between the ages of 18 and 30; neurologically healthy; did not have a history or diagnosis of mental illness; did not use psychoactive medication or drugs; were not color blind; and did not have any permanent metal in their body. Participants were compensated with course credit or €7/h for the behavioral session, and €10/h for the MRI session. The University of Amsterdam Department of Psychology Ethics Committee approved the experiment. All participants signed informed consent beforehand.

Procedure

Participants took part in 2 separate sessions, one behavioral- and one MRI session. During the behavioral session, participants performed a visual search task in which nonspatial auditory stimuli (short tone pips) could be used to facilitate detection of a visual target (a horizontal or vertical line segment) presented among visual distractors (line segments of a different orientation) (see Fig. 1; cf., Van der Burg et al. 2008b). The behavioral session lasted between 45 and 90 min depending on participant's response times. During the MRI session, DTI, MRI, and fMRI auditory/visual localizer data were acquired. Localizer tasks were included that permitted offline localization of primary auditory cortex (A1) and color-sensitive area V4 in the occipital fusiform gyrus (see below). V4 was localized, since in our task, the auditory signal was synchronized to the change in color of the visual target (see next paragraph). These regions of interest were used as seeds in individual probabilistic tractography analyses to reconstruct white-matter pathways connecting these sensory brain regions with their corresponding subcortical structures (see below).

Behavioral Session

The behavioral task used was similar to the one introduced by Van der Burg et al. (2008b), experiment one. It involved searching for a vertical or horizontal target line segment (0.46° visual angle) in a display (7.70° by 7.70° visual angle) of distracter line segments of varying orientation (22.5° deviation from either horizontal or vertical). Participants were asked to indicate whether the target line segment was oriented vertically or horizontally. Prior to the task, participants were explicitly informed about the co-occurrence of the auditory stimulus with the color change of the target stimulus, and encouraged to use the tone to localize the target. An example search display is shown in Figure 1A. Each line segment changed color at a jittered interval (mean: 1.11 Hz; jitter: 50–150 ms). To prevent participants from locating the target immediately after the onset of the search display, target and distracter lines were never presented at fixation. There were 2 set size conditions, determined by the number of
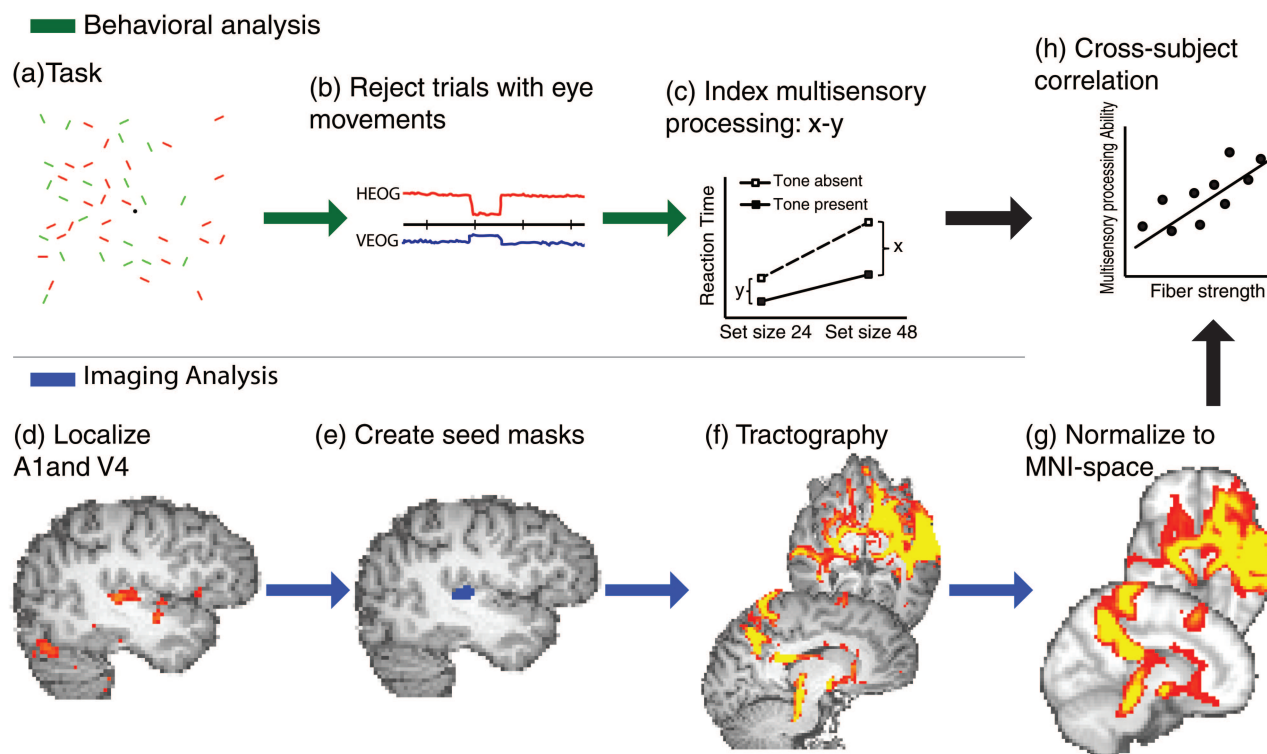
Figure 1. Schematic overview of the main analysis steps. (A) Example of a search display (48 elements) of the audiovisual integration task. The display is shown on a white background for clarity. The target stimulus in the example is located in the lower left-hand side of the panel. (B) Illustration of an EOG trace with eye movement. Trials with eye movements were excluded from the behavioral analysis. (C) To index multisensory processing ability, for each subject, the difference in reaction time between the tone present- and tone absent condition in the small set size condition was subtracted from that of the large set size condition. (D) Sound-related activity based on the auditory localizer data of one representative subject. (E) Left A1 mask based on (D) and used as seed in the tractography analysis. (F) Tractography results of one representative subject in DTI space, and (G) in MNI space. (H) After normalizing all tractography results to MNI space, a cross-subject correlation analysis was run between tract strength and multisensory processing ability separately for each voxel that was connected to the seed region (A1 or V4) in at least 60% of subjects.

distractors surrounding the target: 24 and 48 elements, and a tone-present and tone-absent condition. In the tone-present condition, the color change of the target stimulus coincided with a brief auditory stimulus (duration: 60 ms; frequency: 500 Hz), which contained no information about the orientation, color, or location of the target. In the tone-absent condition, no auditory signal was presented. All conditions were presented randomly intermixed. In total, there were 240 trials (60 per condition).

During the experiment, participants were instructed to keep their eyes fixated on a white dot in the center of the screen, and to covertly search the display for the target stimulus. To ensure participant kept their eyes fixated on the central dot, eye movements were recorded during the task using 4 ocular electrodes. Specifically, vertical eye movements were measured with 2 electrodes placed above and below the left eye (vertical electrooculogram [EOG]). Horizontal eye movements were measured with 2 electrodes placed on the left and right canthi (horizontal EOG). Participants entered their response on a Logitech keyboard and viewed stimuli on a 17″ TFT monitor. They used the "z" and "m" keys to enter their responses. Mapping of these response keys to horizontal or vertical lines was counterbalanced across participants. The auditory stimulus was presented centrally via 2 speakers, with volume kept constant during the experiment. Participants first briefly practiced the task.

MRI Session
Magnetic resonance imaging data were collected on a Philips 3T MRI scanner. DTI data were acquired using single-shot diffusion-weighted spin-echo imaging (ssDWI-SE, 60 slices of 112×112 voxels with a size of $2 \times 2 \times 2$ mm; 32 gradient directions; TR: 7.55 s; TE: 86.16 ms;). To increase DTI signal-to-noise ratio, 4 diffusion-weighted runs with a total acquisition time of ~40 min were collected for each participant. In addition, an anatomical T_1-weighted MRI image (160 slices of 256×256 voxels with a size of $1 \times 1 \times 1$ mm; TR: 8.13 s; TE: 3.72 s) was collected for registration of functional localizer and probabilistic tractography results (see below).

To localize primary auditory cortex and color-sensitive area V4, T_2-*-weighted echo-planar imaging (EPI) images were collected during performance of 2 localizer tasks (flip angle of 76°; 235 and 308 volumes for A1 and V4 localization, respectively; 37 slices of 80×80 voxels with a size of $3 \times 2.5 \times 2.5$ mm and a slice gap of 0.3 mm; TR: 2.0 s; TE: 3.0 s; ascending slice acquisition). The primary auditory cortex was localized using an oddball paradigm. Four tones (440, 550, 660, and 830 Hz) of 250 ms were pseudorandomly presented via scanner-compatible headphones, with an average interstimulus interval (ISI) of 4 s. The ISI was jittered between 2 and 6 s. Participants were instructed to respond by pressing the response button with their right thumb upon hearing the tone with the lowest frequency. To ensure the participants knew which tone was the lowest, it was presented at the beginning of each block. The task consisted of 2 blocks of 40 trials each; with a 3-min break in between.

V4 was localized using a color 1-back task. This localizer task consisted of 16 blocks containing 20 trials, with 6 s of rest in between each block. On each trial, participants were shown a pattern of 9 squares, which were either colored or isoluminant gray, and presented centrally in a 3×3 matrix. Nine colors were used (RGB values: 255,0,0 [red]; 0,0,255 [blue]; 0,255,0 [green]; 128,0,255 [purple]; 255,128,255 [pink]; 255,255,0 [yellow]; 255,128,0 [orange]; 128,64,0 [brown]; 0,128,64 [dark green]). Participants were instructed to press the response button with their right thumb if the composition of the matrix on the present trial was identical to the matrix on the preceding trial (30% of trials). Each block contained only gray squares (control blocks) or colored squares (color blocks). Participants briefly practiced the localizer tasks before going into the MRI scanner.

Behavioral and Eye Movement Data Analysis
EOG data were analyzed using the EEGLAB toolbox in MATLAB (Delorme and Makeig 2004). Trials with eye movements were marked manually and rejected from behavioral analysis (Fig. 1B). Before statistical analysis, on average 22% trials (SD 13%) were removed due to eye movements and nonresponses. Similar to previous studies (Van

der Burg et al. 2008b, 2011), of the remaining trials, the average error rate was 4.8%, and the mean reaction time (RT) for correct trials was 5.3 s, relative to display onset. In addition, error trials and trials in which the participant did not find the target were excluded from RT analysis (cf. Van der Burg et al. 2008b).

We next quantified multisensory processing ability, i.e., the benefit of the auditory tone on visual search efficiency, for each individual separately. Specifically, we calculated the benefit of the tone (tone present [i.e., audiovisual; AV] vs. absent [i.e., visual-only; V]) on visual search efficiency, which is often expressed in terms of the search slope (for a review, see, e.g.,Wolfe and Horowitz 2004) or in our case how much additional time was required for a search in the set size 48 condition versus the set size 24 condition (Fig. 1C). In other words, the difference in RT, obtained by subtracting the RT in the sound-present from the RT in the sound-absent condition in set size 24 condition, was subtracted from the same difference in RT in the set size 48 condition. In equation form: $(V48 - AV48) - (V24 - AV24)$. Importantly, in this double subtraction, pure auditory- and visual-stimulus processing time and processing differences between the set size conditions (e.g., differences related to task difficulty) are subtracted out. The resulting measure has been shown to reflect multisensory processing ability (Van der Burg et al. 2008b, 2011).

Functional Localization and Fiber Tractography
Magnetic resonance imaging data were analyzed using FMRIB Software Library (FSL: www.fmrib.ox.ac.uk/analysis; Woolrich et al. 2001; Smith et al. 2004). First, brains were extracted from functional-, T1-, and DTI data using Brain Extraction Tool (BET; Smith 2002). For each subject, the 4 runs of diffusion-weighted data were concatenated, and individually corrected for motion, angulation, and eddy currents. Following this procedure, the individual gradient directions were rotated per run to account for head motion. From this aggregate DTI image, voxel-specific diffusion tensors were calculated using Bayesian estimation of diffusion parameters obtained using sampling techniques (BEDPOSTX) which uses Markov Chain Monte Carlo sampling to model crossing fibers (Behrens et al. 2007).

To determine whether multisensory processing ability is controlled (in part) by structural connections between sensory-related subcortical structures and auditory and/or visual cortex, we computed probabilistic tractography from primary auditory cortex and color-sensitive area V4, as defined by the fMRI localizers (see Fig. 1D–G). To this end, for each participant separately, a mask was first created for each region of interest (i.e., left and right A1 and V4) based on the sound- and color-related functional localizer activation maps. Functional data were analyzed using the FMRI Expert Analysis Tool (FEAT; Beckmann et al. 2003). The EPI data were corrected for head movement using MCFLIRT motion correction (Jenkinson et al. 2002) and for slice time acquisition. Single participant-level general linear model (GLM) analyses were done using FMRIB's Improved Linear Model (FILM; Woolrich et al. 2001). Time points of auditory stimulus presentation (both target and nontarget), from the auditory localized task, and color/control block onset and offset from the color localized task were used as the explanatory variable and were convolved with a double-gamma hemodynamic response function to model the blood oxygen-level–dependent time course. The temporal derivative was included in the model and motion parameters were added as regressors of no interest.

This localizer analysis revealed brain regions activated by auditory stimuli and color, and the resulted activation maps were used to create participant-specific seed masks in A1 and V4, respectively (see Fig. 1D–E and below). Specifically, first, all individual EPI activation maps were normalized and registered to participant-native DTI space using FMRIB's Linear Image Registration Tool (FLIRT; Jenkinson et al. 2002) with twelve parameter affine registration, and overlaid onto the normalized (to DTI space) T1-weighted anatomical image. Then, for each participant separately, the voxel with the highest peak activity in a predefined region of interest (ROI;, i.e., Heschl's gyrus for A1; and the fusiform gyrus for V4) was manually selected. Forty-nine additional contiguous voxels with the highest sound- or color-related activity were automatically selected, using MATLAB 2011a (The

Mathworks, Inc.), so that the total cluster size of each seed equaled 50 voxels. A mask of these voxels was generated for each hemisphere and participant separately. Next, all masks were manually inspected. Masks where activity extended outside of A1 or V4 (e.g., due to partial volume effects) were manually adjusted. Voxels that extended outside of the ROI were moved to within the ROI to locations where functional activation was next strongest within the ROI. Subsequently, cluster sizes were automatically verified using MATLAB 2011a, to ensure equal cluster sizes across participants. Both medial and lateral parts of Heschl's gyrus were considered A1. In case of 2 Heschl's gyri, both were considered as A1 as well. Often, seeds spanned lateral/medial Heschl's gyrus, and/or included, when present, parts of both Heschl's gyri. Visual area V4 was defined as the occipital fusiform cortex at the approximate height of Lingual gyrus. Fig. 3A shows the average location of the A1 and V4 seeds in MNI space.

To confirm functional localization of A1 and V4, a group average of functional data was created, separately for A1 and V4, by first spatially smoothing (5 mm FWHM) and then normalizing all individual localizer results to MNI space with 6 degrees of freedom, and concatenating them in a 4D volume with participants as the "time" dimension. This group image was used to assess which voxels showed significant activity to auditory stimuli or color at the group level.

To determine probabilistic connectivity between A1 and V4 and other brain structures, participant-specific masks of A1 and V4 were subsequently used as seed voxels in separate probabilistic fiber-tractography analyses using PROBTRACKX, as implemented in FMRIB's diffusion toolbox. For each participant separately, 5000 paths from each voxel in the seed region were generated (cf., Cohen et al. 2008). The curvature threshold was set to 0.2 and step length to 0.5 mm. This procedure resulted in a brain volume for each seed region (i.e., left or right A1 or V4) and participant (Fig. 1F) showing for each voxel the number of received paths. Although this measure (number of paths) is widely used to index the strength or probability of a connection, next to true strength of the underlying pathway, other features such as tract length, fiber geometry, axon density, myelination, and data quality will influence this measure (Johansen-Berg and Rushworth 2009; see also Discussion section). For convenience, however, in the remainder of the article, we use the term "tract strength" to denote the estimated number of paths from a seed region crossing a voxel. Two participants did not show activation in the right auditory cortex and were excluded from the fiber tractography analysis from right A1, making the total number of participants included in the left A1-tractography analysis 25, and that in the right A1-tractography analysis 23. One participant did not show activation in the right V4 and was excluded from fiber tractography analysis from right V4, making the total number of participants included in left V4- tractography analysis 25, and that in the right V4-tractography analysis 24. The resulting "tract strength" volumes were normalized (6 degrees of freedom, i.e., rigid body transformation) and resampled to 3-mm isotropic MNI space (Fig. 1G). This ensured overlap between corresponding brain regions across participants, and reduced the total number of voxels used in the subsequent correlation analysis. Finally, regions in which tracts overlapped with cerebral spinal fluid were masked out, further reducing the number of voxels in the subsequent correlation analysis.

Relationship Between Multisensory Processing Ability and White-Matter Tract Strength

Our main prediction was that estimated strength of anatomical connectivity between auditory and visual cortex and their corresponding subcortical structures would predict individual differences in multisensory processing ability. To test this prediction, we next correlated multisensory processing ability with tract strength indexed by number of paths from the seed crossing each voxel (Fig. 1H). This was done separately for each voxel in which at least 60% of subjects showed non-zero tract strength values (cf. Cohen et al. 2008), and separately for tractography analyses conducted from left A1, right A1, left V4, and right V4.

Statistical Analyses

Behavioral Data

To examine the behavioral effects of a concurrently presented auditory stimulus on visual target detection ability, differences in RTs and errors between conditions of interest were tested with a repeated-measures analysis of variance, using a 2×2 design with set size (24 vs. 48 elements) and tone (present vs. absent) as within participant variables (cf., Van der Burg et al. 2008b experiment 1). The α level was set to 5%.

Functional Localization of A1 and V4

In the analysis of functional localizer data on the individual participant level, the following contrasts were used: Regions involved in processing auditory information were statistically identified by comparing auditory stimulus-related activity to baseline, and regions involved in color processing were identified by comparing color stimulus-related activity to gray stimulus-related activity. All individual functional localization data were tested for significance using t-tests and subsequently z-transformed. Group-level functional localizer data were tested for significance with T-statistics to determine in which voxels activity differed significantly from zero, that is, which voxel showed significant activity to auditory stimuli or color at the group level. Because of our strong a-priori hypotheses (i.e., auditory stimuli activate A1; color stimuli activate V4), and given that the functional localizer analysis was orthogonal to the tractography analysis, no cluster correction and uncorrected ($\alpha = 5\%$) P-values were used.

Relationship Between Tract Strength and Multisensory Processing Ability

To assess the relationship between tract strength and multisensory processing ability across subjects, Spearman's rank correlation coefficient was used. Rank correlations are less sensitive to outliers and violations of the assumption of normally distributed data. For statistical thresholding of correlation maps, a 2-step nonparametric permutation approach (Nichols and Holmes 2002) was used ($P < 0.005$ and contiguous cluster threshold of minimally 15 voxels; see below). At the first stage (voxel level), the assignment of behavioral data to brain data was shuffled, and voxelwise correlations were computed iteratively 1000 times. The resulting correlation distribution was converted to a z-distribution, and the standardized value of the correlation coefficient of nonshuffled behavioral data was computed per voxel. To correct for multiple comparisons, the standardized correlation map was thresholded at z-scores corresponding to P-values of 0.005. At the second stage (cluster level), a distribution of maximum cluster sizes was computed under the null hypothesis, by correlating shuffled behavioral data with tract strength at each voxel for 1000 iterations. The cluster size corresponding to the 95th percentile of the resulting maximum cluster size distribution was taken as the lower limit for cluster thresholding of the standardized correlation map (Nichols and Holmes 2002). Cluster thresholds were 15 and 16 voxels for right and left hemisphere, respectively, for A1, and 15 for both hemispheres for V4. Previous DTI studies have followed a similar thresholding procedure (e.g., Cohen et al. 2008).
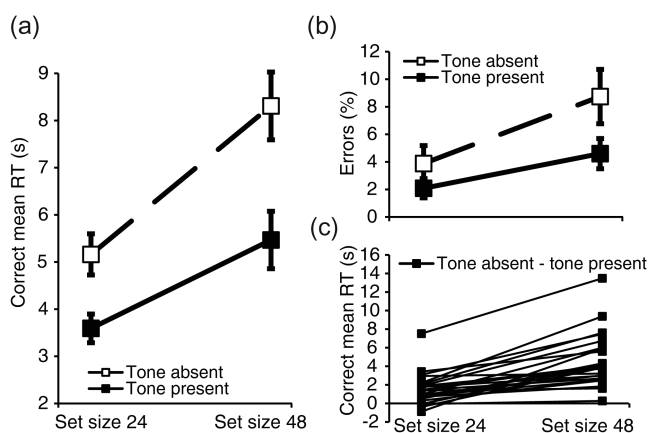
Post hoc Control Analyses

In order to demonstrate the specificity of our results (see below) and to exclude potential confounds, such as age, gender (Tomasi et al. 2008), local tissue properties unrelated to connectivity (Cohen 2011; de Wit et al. 2012) and cerebral volume, we ran several post hoc control analyses. These are described in the Supplementary Material.

Results

Behavior

In the audiovisual integration task, the visual target stimulus was either surrounded by 47 distractors (high-cluttered trials), or by 23 distractors (low cluttered trials) (see Fig. 1). As expected, in the high cluttered trials (set size 48), participants were slower and less accurate to detect the visual target than in the low cluttered trials (set size 24; see Fig. 2A,B). This difference (or search efficiency) was reflected by a main effect of set size for RT: $F_{1,24} = 63.58$, $P < 0.001$; and for response accuracy: $F_{1,24} = 15.03$, $P = 0.001$. Importantly, the auditory signal improved both the speed of detection of the synchronized visual target (main effect of tone for RT: $F_{1,24} = 27.64$, $P < 0.001$), and response accuracy ($F_{1,24} = 7.63$, $P = 0.01$). Moreover, tone-related improvements in search time were particularly pronounced in the highly cluttered trials (significant interaction between tone and set size, $F_{1,24} = 10.73$, $P = 0.003$), indicating that the nonspatial tone improved visual search efficiency. Notably, as in previous studies using a highly similar task (Van der Burg et al. 2008a, 2008b), there was large individual variability in the extent to which an individual's visual search efficiency was improved by the presence of the synchronized auditory signal (or multisensory processing ability) (Fig. 2C). In the current study, the size of this multisensory effect varied between −2.18 and 5.58 s across individuals—with larger positive values denoting a greater benefit of the auditory signal on search efficiency (or greater multisensory processing ability). No interaction between tone and set size was present for response accuracy ($F_{1,24} = 2.83$, $P = 0.105$).

Functional Localization of Sensory Cortical Areas

A1 and V4 were successfully localized as indicated by the group-localizer results (see Fig. 3B for group-average A1 and V4 activity); bilateral clusters of auditory-related activation were found in left and right A1 (peak t-values of 8.46 [xyz MNI coordinates: $x = −39$; $y = −21$ $z = 6$] and 7.83 [$x = 48$; $y = −24$; $z = 12$], respectively). Bilateral clusters of color-related activation were found in left and right V4 (peak t-values of 7.91 [xyz MNI coordinates: $x = −24$; $y = −72$ $z = −6$] and 7.92 [$x = 24$; $y = −75$; $z = −9$], respectively).

Relationship Between Subcortical–Cortical Connectivity and Multisensory Processing

On average, per voxel included in our tractography analyses, the mean number of participants showing connectivity with A1 was 18.26 and 17.62 for the left and right A1 seeds, respectively. On average, per included voxel, the mean number of participants showing connectivity with V4 was 18.73 and 18.86 for the left and right V4 seeds, respectively.

Our main prediction was that estimated strength of anatomical connectivity between auditory and visual cortex and their corresponding subcortical structures would predict individual differences in multisensory processing ability. In line with our prediction, participants with better multisensory processing ability had stronger connections (as indicated by the number of paths) between left A1 and left auditory thalamus (MGN; MNI coordinates of peak z-statistic: $x = −6$ mm; $y = −30$ mm; $z = +3$ mm), and between left A1 and left cochlear nucleus (MNI coordinates of peak z-statistic: $x = −9$ mm; $y = −39$ mm; $z = −39$ mm) (see Fig. 3B). These findings provide support for the idea that in humans, low-level connections within putatively modality-specific pathways contribute to multisensory processing. Strength of the identified pathways between V4 and the thalamus did not predict multisensory processing ability. All clusters of correlation are shown in Table 1.

Of importance, a post hoc control tractography analysis with the parietal operculum (S2) (a brain region neighboring A1) as the seed did not reveal a significant correlation between tract strength to S2 and multisensory processing ability near the MGN or the cochlear nucleus (see

Figure 2. Behavioral results. Effects of the synchronized auditory tone on visual target detection times (A) and accuracy (B) in the 2 set size conditions. Replicating previous studies, the auditory signal greatly reduced visual target search times in the high vs. low set size condition. Yet, notably, there were large individual differences in the size of this benefit (C). The percentage of error trials is based only on those trials that did not contain eye movements. Error bars represent the standard error of the mean (SEM).

Table 1.

Brain regions in which strength of connectivity to A1 or V4 predicted multisensory processing ability across participants

Seed	Location	MNI coordinates (x y z)	Size (n voxels)	Peak z-statistic
Left A1	POC	−36, −33, 27	41	3.46
	LOC	−39, −60, 15	23	3.49
	aTh	−12, −9, 6	81	3.20
	Put.	−24, −3, 3	81	3.64
	Amyg.	−18, −12, −12	76	3.22
	Cer.	−15, −60, −30	41	3.51
	MGN	−6, −27, 6	37	3.10
	CN	−9, −39, −39	17	2.77
Right A1	AG	42, −54, 42	47	3.25
	LOC	45, −63, −6	82	3.69
	SCC	60, −12	58	3.11
Left V4	FFG	−42, −21, −27	20	−3.37

Location of correlations was assessed using the Harvard–Oxford structural atlas. Peak MNI coordinates are indicated in mm. All clusters were found in their respective seed hemisphere. All correlations seeded from A1 were positive, and all correlations seeded from V4 were negative. POC, parietal operculum cortex; LOC, lateral occipital cortex; aTh, anterior thalamus; Put., putamen; Amyg.: amygdala; Cer.: cerebellum; MGN: medial geniculate nucleus; CN, cochlear nucleus; AG: angular gyrus; SCC, subcallosal cortex; FFG, fusiform gyrus.
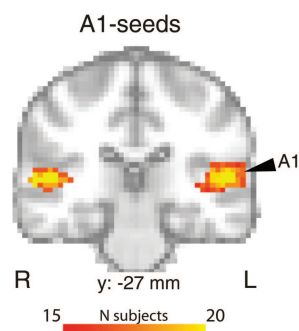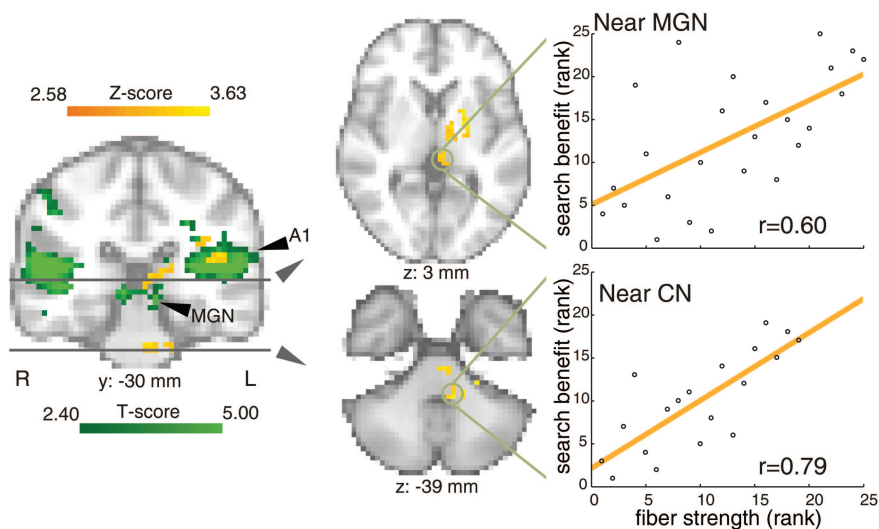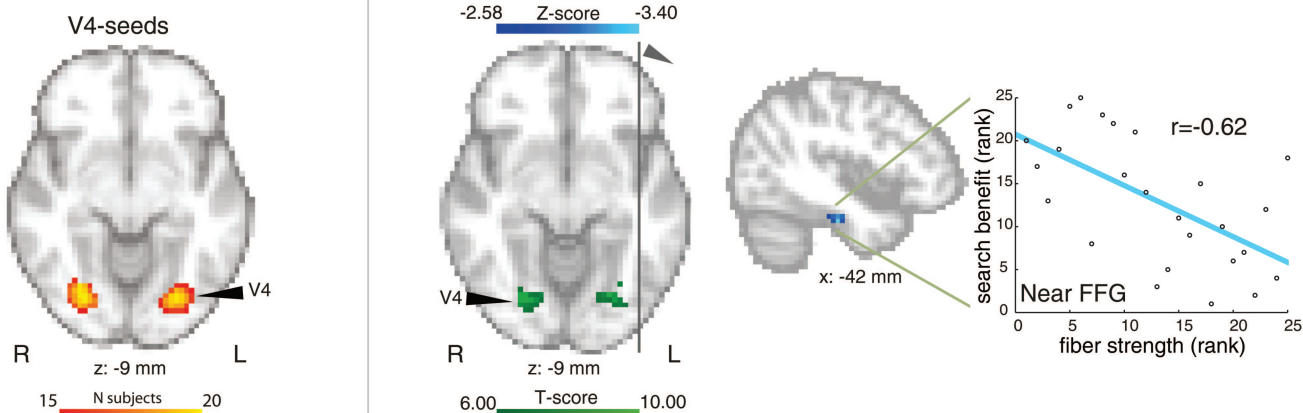
Figure 3. Low-level connections within the auditory system govern multisensory processing ability. (A) Average location of individual A1 and V4 seeds, normalized to MNI space. Colors indicate the number of subjects for which the seed overlapped within a given voxel. (B) Voxels in which fiber tract strength seeded from A1-predicted multisensory processing ability are shown in red–yellow. Group-level functional activation results of the auditory localizer task are also shown (in green) to indicate the location of A1 and MGN. As predicted, multisensory processing ability correlated with the strength of white-matter tracts between primary auditory cortex, MGN, and the cochlear nucleus, the first relay station in the auditory processing stream. (C) Identified pathways between V4 and the thalamus were not associated with multisensory processing ability. A1, primary auditory cortex; MGN, medial geniculate nucleus; CN, cochlear nucleus; FFG, fusiform gyrus; R, right; L, left.

Supplementary Material). This control analysis highlights the specificity of our findings and supports the idea that anatomical connections within the auditory system contribute to multisensory processing ability.

Additionally, a second post hoc analysis based on the auditory functional localizer data (see Supplementary Material) revealed sound-related activity in both the left and right MGN (xyz MNI coordinates: [$x = -9$; $y = -30$ $z = -6$] and [$x = 9$; $y = -30$; $z = -3$], respectively; cf., (Devlin et al. 2006; Noesselt et al. 2010). Importantly, as can be seen in Figure 3B, the voxels which displayed a correlation between tract strength to A1 and multisensory processing ability in our original analysis were located slightly more dorsal from where MGN was localized functionally in our study as well as in previous studies (Devlin et al. 2006; Noesselt et al. 2010), and seem to be part of a tract that enters MGN. This may indicate that connection strength within voxels that are part of a tract connecting A1

and MGN—rather than voxels within MGN itself—predicted multisensory processing ability.

Lastly, we examined the relationship between multisensory processing ability and local tissue properties, including white-matter integrity (or fractional anisotropy [FA]), mean diffusivity (MD), and gray matter density in those voxels in which tract strength determined multisensory processing ability (see Methods section in Supplementary Material). No significant correlations were observed (Supplementary Fig. S1 and S2). Because FA, MD, and gray matter density reflect only local tissue properties at each voxel and are unrelated to connectivity with the seed region, these results confirm the specificity of the correlations shown in Figure 3 to specific paths linking, for example, A1 to MGN and cochlear nucleus. Of further note, in the voxels in which tract strength predicted multisensory processing ability, mean FA and MD were relatively high, suggestive of underlying white matter (mean FA

in significant voxels seeded from left A1 and V4, respectively: 0.52 and 0.37. Mean MD in significant voxels seeded from left A1 and V4, respectively: 5.06·10−4 and 3.80·10−4. compare to e.g., Cohen 2011).

Discussion

The current study tested the hypothesis that subcortical pathways, which traditionally have been considered modality-specific, contribute to multisensory processing in humans using an audiovisual integration task and DTI. In line with our hypothesis, we found that the strength of anatomical connections between the first relay system in the auditory processing stream (i.e., the cochlear nucleus in the brain stem), the auditory thalamus, and the primary auditory cortex predicted one's ability to combine auditory and visual information. This novel finding indicates that putatively modality-specific subcortical structures may contribute to multisensory processing in humans by way of their (reciprocal) connections with sensory cortices. It corroborates previous work in animals (e. g., Benevento and Fallon 1975; Itaya and Van Hoesen 1982; Linke 1999; Malmierca et al. 2002), which has demonstrated reciprocal connections at early, subcortical processing stages between modality-specific systems. Our finding also extends recent findings from functional neuroimaging studies in humans (Baier et al. 2006; Musacchia et al. 2006; Noesselt et al. 2010), which have implicated subcortical processing stages in multisensory processing in humans, by showing that strength of subcortical–cortical anatomical connections predicts the efficacy of multisensory processing. Together these findings suggest integration mechanisms that go beyond traditional models based on a hierarchical convergence of sensory processing. To our knowledge, the present study is the first to indicate that multisensory processing ability is reflected in anatomical connections in the human brain.

An important question is how connections between subcortical structures, such as the thalamus, and sensory cortices may contribute to multisensory processing. One candidate mechanism is the cross-modal phase resetting of cortical activity in one modality by input from another modality. Recent research in monkeys (Lakatos et al. 2007; Kayser et al. 2008) and humans (Naue et al. 2011; Thorne et al. 2011) suggests that through phase resetting of oscillatory activity in one modality (e.g., visual), stimulation in another modality (e. g., auditory) may produce perceptual amplification of input in that modality (i.e., visual). Notably, it has been proposed that the thalamus is uniquely important in promoting phase resetting of ongoing oscillatory activity and cortical synchrony (Lakatos et al. 2007). Specifically, it has been postulated that oscillatory input from thalamic cells in primary sensory relay nuclei such as the MGN to the cortex can promote widespread synchronization of activity between other thalamic nuclei via corticothalamic feedback projections, and through them, other cortical regions. This way, the thalamus can establish synchrony of cortical oscillatory activity in different modalities, and hence, contribute to multisensory interactions at the level of primary sensory cortex. An intriguing possibility warranting future research combining DTI and electroencephalography (EEG) is thus that anatomical connections between thalamus and sensory cortices implement the multisensory effects which have been observed at very short delays in sensory cortices in recent EEG and ERP studies (Molholm

et al. 2002; Naue et al. 2011; Thorne et al. 2011; Van der Burg et al. 2011).

Interestingly, a cross-participant relationship between audiovisual processing ability and anatomical connection strength was only observed between brain areas within the auditory system, but not within pathways that were tracked from a region neighboring A1 (i.e., the parietal operculum) or within pathways that were tracked from V4 to subcortical structures (including LGN), highlighting the specificity of our finding. As our task measured one's ability to benefit from auditory information during a difficult visual search task, it is well possible that studies using a task that measures the added benefit of visual information on auditory processing (e.g., Busse et al. 2005) would find the opposite result. As in our task, the auditory signal was synchronized with the color change of the visual target, color area V4 was used as a seed for probabilistic tractography. It is also possible that a seed in another visual area would have provided different results. Additionally, as our sample consisted largely of female subjects, future research is also necessary to examine to what extent gender differences in brain function (e.g., Tomasi et al. 2008) contribute to the observed relationship between tract strength and multisensory processing ability. We should note that excluding the male subjects (see Supplementary Information) from our analysis did not alter the pattern of results.

It is noteworthy, as noted in the introduction, that animal research has revealed a direct, short-latency projection between the cochlear nucleus and MGN (Malmierca et al. 2002; Anderson et al. 2006), which parallels the classical lemniscal auditory pathway involved in the perception of sound. This direct cochlear nucleus-MGN (or nonlemniscal) pathway links auditory input with inputs from visuomotor and other nonauditory systems, and is hence thought to contribute to multisensory integration. Our correlation between multisensory processing ability and tract strength to A1 in pathways connecting the cochlear nucleus and MGN to A1 may thus well reflect individual differences in multisensory processing rather than individual differences in the ability to process auditory information per se. Yet, it should be noted that we only obtained estimates of connectivity with A1 within the auditory system, not of connectivity between the cochlear nucleus and MGN.

As it is not possible to determine directionality of anatomical connections based on DTI data, the observed correlations between tract strength and multisensory processing ability may reflect enhanced connectivity from subcortical structures to auditory cortex (bottom-up) or vice versa (top-down) or both. Histological tracing studies in animals have shown bottom-up connectivity between modality-specific pathways (e.g., Malmierca et al. 2002), however, and it is therefore conceivable that multisensory processing occurs at early, subcortical processing stages. Next to the issue of polarity, DTI has several other limitations for tractography. For example, it is known to perform less well in brain areas of high fiber curvature or with crossing or kissing fibers. Despite these limitations, findings from studies using physical phantoms, animal models, or postmortem human brains indicate that diffusion can principally model the underlying microstructures and white-matter pathways even on a small spatial scale. Of particular importance, a recent study found that tractography results map well onto real white-matter pathways in the human brain (Seehaus et al. 2013), supporting the common

assumption in this and most prior DTI-tractography studies (e.g., Cohen et al. 2009) that DTI tractography captures white-matter pathways. Moreover, we were able to reconstruct known anatomical pathways within the auditory system that have been identified in histological tracing studies. A previous DTI study in humans successfully tracked these small-scale white-matter connections within the auditory system as well (Devlin et al. 2006). Finally, our findings agree with independent data from a recent fMRI study (Noesselt et al. 2010), showing that functional connectivity between the MGN and primary auditory cortex is modulated by multisensory processing and predicts multisensory task performance. Together, the above strengthens an interpretation of the observed correlation between multisensory processing ability and tract probability in terms of individual differences in the strength of connections within the putatively auditory system.

In summary, the current study shows that one's ability to combine auditory and visual information during a visual search task is predicted by the strength of connections between the first relay station within the auditory pathway, the auditory thalamus, and the primary auditory cortex. This finding adds to a growing body of research that indicates that subcortical sensory pathways do not only feed forward unisensory information to the cortex, and suggests that multisensory processing is governed by low-level brain anatomy in humans.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxford-journals.org/.

Funding

Notes

References

Anderson LA, Malmierca MS, Wallace MN, Palmer AR. 2006. Evidence for a direct, short latency projection from the dorsal cochlear nucleus to the auditory thalamus in the guinea pig. Eur J Neurosci. 24:491–498.

Baier B, Kleinschmidt A, Muller NG. 2006. Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. J Neurosci. 26:12260–12265.

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.

Beckmann CF, Jenkinson M, Smith SM. 2003. General multilevel linear modeling for group analysis in FMRI. Neuroimage. 20:1052–1063.

Behrens TEJ, Berg HJ, Jbabdi S, Rushworth MFS, Woolrich MW. 2007. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage. 34:144–155.

Benevento LL, Fallon JH. 1975. The Ascending projections of the Superior Colliculus in the Rhesus monkey (macaca mulatta). J Comp Neurol. 160.

Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG. 2005. The spread of attention across modalities and space in a multisensory object. Proc Natl Acad Sci U S A. 102:18751–18756.

Cappe C, Rouiller EM, Barone P. 2009. Multisensory anatomical pathways. Hear Res. 258:28–36.

Cohen MX. 2011. Hippocampal-prefrontal connectivity predicts midfrontal oscillations and long-term memory performance. Curr Biol. 21:1900–1905.

Cohen MX, Elger CE, Weber B. 2008. Amygdala tractography predicts functional connectivity and learning during feedback-guided decision-making. Neuroimage. 39:1396–1407.

Cohen MX, Schoene-Bake JC, Elger CE, Weber B. 2009. Connectivity-based segregation of the human striatum predicts personality characteristics. Nat Neurosci. 12:32–34.

Delorme A, Makeig S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 134:9–21.

Devlin JT, Sillery EL, Hall DA, Hobden P, Behrens TEJ, Nunes RG, Clare S, Matthews PM, Moore DR, Johansen-Berg H. 2006. Reliable identification of the auditory thalamus using multi-modal structural analyses. Neuroimage. 30:1112–1120.

de Wit S, Watson P, Harsay HA, Cohen MX, van de Vijver I, Ridderinkhof KR. 2012. Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. J Neurosci. 32:12066–12075.

Driver J, Noesselt T. 2008. Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. Neuron. 57:11–23.

Fairhall SL, Macaluso E. 2009. Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. Eur J Neurosci. 29:1247–1257.

Fort A, Delpuech C, Pernier J, Giard MH. 2002. Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. Cereb Cortex. 12:1031–1039.

Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. 2005. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J Neurosci. 25:5004–5012.

Ghazanfar AA, Schroeder CE. 2006. Is neocortex essentially multisensory? Trends Cogn Sci. 10:278–285.

Giard MH, Péronnet F. 1999. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci. 11:473–490.

Itaya SK, Van Hoesen GW. 1982. Retinal innervation of the inferior colliculus in rat and monkey. Brain Res. 233:45–52.

Jenkinson M, Bannister P, Brady M, Smith S. 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. 17:825–841.

Johansen-Berg H, Rushworth MF. 2009. Using diffusion imaging to study human connectional anatomy. Annu Rev Neurosci. 32:75–94.

Jones EG. 2007. The Thalamus. Cambridge (MA): Cambrigde University Press.

Kayser C, Logothetis NK. 2007. Do early sensory cortices integrate cross-modal information? Brain Struct Funct. 212:121–132.

Kayser C, Petkov CI, Logothetis NK. 2008. Visual modulation of neurons in auditory cortex. Cereb Cortex. 18:1560–1574.

Komura Y, Tamura R, Uwano T, Nishijo H, Ono T. 2005. Auditory thalamus integrates visual inputs into behavioral gains. Nat Neurosci. 8:1203–1209.

Lakatos P, Chen C-M, O'Connell MN, Mills A, Schroeder CE. 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron. 53:279–292.

Linke R. 1999. Differential projection patterns of superior and inferior colliculus neurons onto posterior paralaminar nuclei of the thalamus surrounding the medial geniculate body in the rat. Eur J Neurosci. 11:187–203.

Macaluso E. 2006. Multisensory processing in sensory-specific cortical areas. Neuroscientist. 12:327–338.

Malmierca MS, Merchán MA, Henkel CK, Oliver DL. 2002. Direct projections from cochlear nuclear complex to auditory thalamus in the rat. J Neurosci. 22:10891–10897.

Martuzzi R, Murray MM, Michel CM, Thiran JP, Maeder PP, Clarke S, Meuli RA. 2006. Multisensory interactions within human primary cortices revealed by BOLD dynamics. Cereb Cortex. 17:1672–1679.

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. 2002. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cogn Brain Res. 14:115–128.

Mori S, Zhang J. 2006. Principles of diffusion tensor imaging and its applications to basic neuroscience research. Neuron. 51:527–539.

Musacchia G, Sams M, Nicol T, Kraus N. 2006. Seeing speech affects acoustic information processing in the human brainstem. Exp Brain Res. 168:1–10.

Naue N, Rach S, Struber D, Huster RJ, Zaehle T, Korner U, Herrmann CS. 2011. Auditory event-related response in visual cortex modulates subsequent visual responses in humans. J Neurosci. 31:7729–7736.

Nichols TE, Holmes AP. 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp. 15:1–25.

Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J. 2007. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. J Neurosci. 27:11431–11441.

Noesselt T, Tyll S, Boehler CN, Budinger E, Heinze HJ, Driver J. 2010. Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. J Neurosci. 30:13609–13623.

Schroeder CE, Foxe J. 2005. Multisensory contributions to low-level, "unisensory" processing. Curr Opin Neurobiol. 15:454–458.

Schroeder CE, Foxe JJ. 2002. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cogn Brain Res. 14:187–198.

Seehaus AK, Roebroeck A, Chiry O, Kim DS, Ronen I, Bratzke H, Goebel R, Galuske RAW. 2013. Histological validation of DW-MRI tractography in human postmortem tissue. Cereb Cortex. 23(2):442–450.

Smith SM. 2002. Fast robust automated brain extraction. Hum Brain Mapp. 17:143–155.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE et al. 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 23:S208–S219.

Stein BE, Stanford TR. 2008. Multisensory integration: Current issues from the perspective of the single neuron. Nat Rev Neurosci. 9:255–266.

Thorne JD, De Vos M, Viola FC, Debener S. 2011. Cross-modal phase reset predicts auditory task performance in humans. J Neurosci. 31:3853–3861.

Tomasi D, Chang L, Caparelli EC, Ernst T. 2008. Sex differences in sensory gating of the thalamus during auditory interference of visual attention tasks. Neuroscience. 151:1006–1015.

van Atteveldt NM, Blau VC, Blomert L, Goebel R. 2010. fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. BMC Neurosci. 11:1–11.

van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J. 2008a. Audiovisual events capture attention: evidence from temporal order judgments. J Vis. 8(5):1–10.

van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J. 2008b. Pip and pop: nonspatial auditory signals improve spatial visual search. J Exp Psychol Hum Percept Perform. 34:1053–1065.

van der Burg E, Talsma D, Olivers CNL, Hickey C, Theeuwes J. 2011. Early multisensory interactions affect the competition among multiple visual objects. Neuroimage. 55:1208–1218.

Wolfe JM, Horowitz TS. 2004. What attributes guide the deployment of visual attention and how do they do it? Nat Rev Neurosci. 5:1–7.

Woolrich MW, Ripley BD, Brady M, Smith SM. 2001. Temporal autocorrelation in univariate linear modeling of fMRI data. Neuroimage. 14:1370–1386.