Behavioral/Systems/Cognitive

Distinct Functional Contributions of Primary Sensory and Association Areas to Audiovisual Integration in Object Categorization

Sebastian Werner and Uta Noppeney

Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

Multisensory interactions have been demonstrated in a distributed neural system encompassing primary sensory and higher-order association areas. However, their distinct functional roles in multisensory integration remain unclear. This functional magnetic resonance imaging study dissociated the functional contributions of three cortical levels to multisensory integration in object categorization. Subjects actively categorized or passively perceived noisy auditory and visual signals emanating from everyday actions with objects. The experiment included two 2 × 2 factorial designs that manipulated either (1) the presence/absence or (2) the informativeness of the sensory inputs. These experimental manipulations revealed three patterns of audiovisual interactions. (1) In primary auditory cortices (PACs), a concurrent visual input increased the stimulus salience by amplifying the auditory response regardless of task-context. Effective connectivity analyses demonstrated that this automatic response amplification is mediated via both direct and indirect [via superior temporal sulcus (STS)] connectivity to visual cortices. (2) In STS and intraparietal sulcus (IPS), audiovisual interactions sustained the integration of higher-order object features and predicted subjects' audiovisual benefits in object categorization. (3) In the left ventrolateral prefrontal cortex (vlPFC), explicit semantic categorization resulted in suppressive audiovisual interactions as an index for multisensory facilitation of semantic retrieval and response selection. In conclusion, multisensory integration emerges at multiple processing stages within the cortical hierarchy. The distinct profiles of audiovisual interactions dissociate audiovisual salience effects in PACs, formation of object representations in STS/IPS and audiovisual facilitation of semantic categorization in vlPFC. Furthermore, in STS/IPS, the profiles of audiovisual interactions were behaviorally relevant and predicted subjects' multisensory benefits in performance accuracy.

Introduction

To enable effective perception and action in our multisensory environment, the human brain merges information from multiple senses. Behaviorally, multisensory integration improves detection, discrimination and categorization (Calvert et al., 2004). While multisensory integration was conventionally thought to be deferred until later processing stages in association cortices (Calvert, 2001), recent studies have shown multisensory influences in primary, putatively unisensory regions (Schroeder and Foxe, 2002; Foxe and Schroeder, 2005). Provocatively, the entire neocortex has been defined as "multisensory" (Ghazanfar and Schroeder, 2006). This multitude of integration sites requires us to move beyond simply designating brain areas as multisensory toward characterizing the functional similarities, differences and constraints that govern multisensory processes at different cortical levels. Coarsely, three processing stages may be dissociated where multisensory influences emerge during object categorization. First, multisensory costimulation within a narrow spatiotemporal window increases stimulus salience. Second, higher-order features are extracted and integrated into object representations. Third, semantic retrieval enables object categorization and selection of an appropriate action. Previous electrophysiological and functional imaging studies have characterized multisensory properties of brain regions primarily by manipulating spatial (where?) (Wallace et al., 1996; Macaluso and Driver, 2005), temporal (when?) (Calvert et al., 2000; Noesselt et al., 2007; van Atteveldt et al., 2007; Kayser et al., 2008) and semantic (what?) (Hein et al., 2007; Noppeney et al., 2008) congruency of signals from different senses. However, incongruency manipulations violate natural multisensory relationships and invoke error detection processes. Hence, their role in characterization of natural multisensory integration processes may be limited.

To dissociate the neural processes underlying multisensory (1) salience effects due to costimulation, (2) formation of object percepts and (3) facilitation of semantic categorization and response selection, the present study manipulates the audiovisual input, subjects' behavioral performance and task. In all experimental conditions, subjects were presented with noisy dynamic auditory and/or visual signals emanating from everyday object actions. Subjects actively categorized the objects or passively attended to them while involved in a target detection task. Cru-

Received Oct. 13, 2009; revised Dec. 10, 2009; accepted Dec. 12, 2009.

This work was supported by Deutsche Forschungsgemeinschaft and the Max Planck Society. We thank Richard Lewis for his very helpful comments on a previous version of this manuscript and Mario Kleiner for help with stimulus preparation.

Correspondence should be addressed to Sebastian Werner, Max Planck Institute for Biological Cybernetics, Spemannstrasse 41, 72076 Tübingen, Germany. E-mail: sebastian.werner@tuebingen.mpg.de.

DOI:10.1523/JNEUROSCI.5091-09.2010

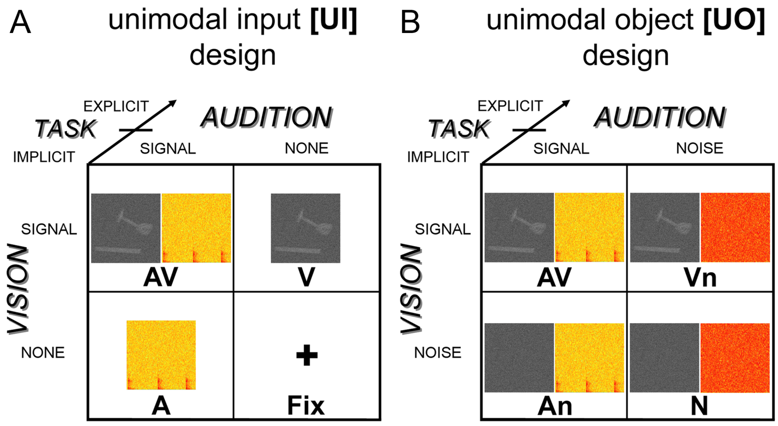

Figure 1. Experimental design and example stimuli. ***A***, Unimodal input design manipulating: (1) visual input: presence vs absence; (2) auditory input: presence vs absence; (3) task: explicit vs implicit. ***B***, Unimodal object design manipulating: (1) visual informativeness: object information vs noise; (2) auditory informativeness: object information vs noise; (3) task: explicit vs implicit. Example stimulus: ***A***, Hammer as one frame of the video clip and the spectrogram (0–2.5 kHz, 2 s) of the corresponding source sound. ***B***, Same stimuli as in ***A*** but here visual object stimuli were presented with an auditory white noise (Vn) and auditory object stimuli with a dynamic visual white noise (An). Fix was replaced by N, a dynamic audiovisual white noise.

cially, the experiment included two 2 × 2 factorial designs that enabled the computation of audiovisual interactions at two levels. (1) The Unimodal Input design [UI] manipulated the absence/presence of the auditory (resp. visual) input. Here, audiovisual interactions can emerge due to both, costimulation per se and integration of higher-order object features. (2) The Unimodal Object information [UO] design provided low-level auditory and visual inputs in all conditions but manipulated their informativeness by adding different amounts of noise. Therefore, the [UO] design controls for effects of costimulation and selectively focuses on integration of object information. We used the following rationale to dissociate the neural processes underlying these three processing stages. (1) The effect of "costimulation" was revealed by comparing the audiovisual interactions of the [UI] relative to those of the [UO] design. (2) Regions associated with the formation of an object percept were identified by relating subjects' audiovisual interactions to their multisensory benefits in object categorization. (3) Audiovisual categorization processes were expected to be enhanced during explicit categorization relative to passive (implicit) stimulus exposure (Fig. 1).

Materials and Methods

Subjects

Twenty-one right-handed subjects (10 females; mean age: 24.2; SD: 1.9) with no history of neurological or psychiatric illness gave informed consent to participate in the study. All subjects had normal or corrected to normal vision and reported normal hearing. The study was approved by the human research ethics committee of the medical faculty at the University Tübingen.

Stimuli

Audiovisual movies of actions performed with 15 tools (e.g., hammer, saw, drill, scissors) and 15 musical instruments (e.g., drum, guitar, flute, violin) were selected to enable a semantic categorization task. Yet, category-selective activations (Chao et al., 1999; Lewis et al., 2005; Noppeney et al., 2006) are not the focus of this communication. To increase the probability of superadditive interactions, both auditory and visual components were presented in a degraded manner (Meredith and Stein, 1986; Stanford et al., 2005; Stevenson and James, 2009).

Visual object stimuli were gray-scale video clips (50 frames at 25 fps, 2 s duration, size 8.5° × 10.5° visual angle) recorded at the Max Planck Institute-VideoLab (Kleiner et al., 2004). The video clips were degraded by weighted averaging the original movie frames with random noise images of the identical size. Auditory object stimuli were sounds produced by the actions. Each sound clip (2 s duration, 48 kHz sampling rate, presented at ~84 dB SPL) was equated for maximum intensity of the sound stimulation. Similar to the procedure for degrading visual images, auditory stimuli were a weighted average of the original sound clip with a random noise sound of identical length. The item-specific degradation level was determined based on a behavioral pilot study to obtain an across-subject categorization accuracy of 75% averaged over all items within a category (see supplemental materials, available at www.jneurosci.org as supplemental material). To prevent subjects from using low-level visual or auditory cues for categorization, the audiovisual movies of the two categories (i.e., tools or musical instruments) were matched with respect to their mean luminance (t_{28} = 0.29; $p > 0.05$), root-mean-square (RMS) contrast ($t_{28} = 0.91$; $p > 0.05$) and their auditory RMS power ($t_{28} = 1.86$; $p > 0.05$).

During the target detection task (only), simple visual (centrally presented mid-level gray circle on black background; 4° visual angle), auditory (complex tone produced by combining five sinusoidal components at 440, 880, 1760, 3520 and 7040 Hz frequencies with the power of components exponentially decreasing with frequency) or audiovisual (circle+tone) targets were presented interspersed among the object stimuli (i.e., tools and musical instruments). The target stimuli were presented for 300 ms.

Experimental design

The experiment included two 2 × 2 factorial designs that enable the identification of multisensory integration sites through the interaction between the visual and the auditory factors (Calvert, 2001) (Fig. 1): (1) The unimodal input [UI] design manipulated the presence/absence of visual and auditory inputs. Here, the interaction terms encompass the effect of audiovisual costimulation and integration of object information (Fig. 1*A*). (2) The unimodal object [UO] design manipulates the informativeness of the sensory inputs (i.e., object information vs noise). In this design, low-level auditory and visual inputs are present in all conditions. Hence, the interaction term of this design controls for effects of costimulation and selectively focuses on the integration of higher-order object information (Fig. 1*B*). Directly comparing the interaction terms of the [UI] and the [UO] design ([UI]>[UO]) should thus selectively reveal interactions due to low-level effects of audiovisual costimulation and associated salience effects (see supplemental Table S4, available at www.jneurosci.org as supplemental material, for analysis rationale).

In addition, we manipulated task-context in both designs (note: this additional task manipulation turns both 2 × 2 [UI] and [UO] designs in two 2 × 2 × 2 factorial designs; Fig. 1). Subjects either categorized the object stimuli as tools or musical instruments (the EXPLICIT task) or passively perceived them while engaged in a target detection task (the IMPLICIT task). During the target detection task, subjects responded to simple visual, auditory and audiovisual targets that were presented randomly interspersed in the stream of object stimuli. Since these target stimuli were presented only during the target detection task and not during the categorization task, they were modeled in the functional magnetic resonance imaging (fMRI) analysis separately from the main experimental stimuli of interest to render the stimulus-induced activations comparable across task contexts. Yet, since these additional targets were interspersed (i.e., not co-occurring) with the main experimental stimuli (tools and instruments), subjects may have systematically switched attention between the two multisensory streams (i.e., the task-relevant targets and the irrelevant tools/instruments stimuli) in the IMPLICIT context only. The target detection task was designed to render the "implicit" conditions comparable to previous passive viewing/listening paradigms while ensuring a minimum level of control on subjects' attentional level. The task-manipulation enables us to dissociate automatic and categorization-related audiovisual integration processes.

Experimental procedure

Video clips and sound stimuli were presented for 2000 ms followed by 800 ms fixation. Each of the 30 items (15 tools and 15 musical instruments) was presented twice in each condition. During the (explicit) categorization task, subjects categorized auditory, visual and audiovisual stimuli as tools or musical instruments as quickly and accurately as possible via a two choice keypress. During the target detection task, they responded as fast as possible to simple visual (circle), auditory (tone) or audiovisual (circle+tone) targets, while passively (i.e., implicitly) perceiving the object stimuli (i.e., tools and musical instruments). Approximately, 15% of the trials were targets. Object trials (i.e., AV, V, Vn, A, An), audiovisual noise (i.e., N) trials and null events (i.e., Fix trials) were the 7 conditions later modeled in the fMRI analysis. Task instructions were given at the beginning and in the middle of each scanning session via visual display (i.e., a continuous half of each scanning session was dedicated to a single task). The stimuli were presented in blocks of 8 stimuli. The stimulus blocks were interleaved with 6 s fixation. During the detection task, each block contained at least one target, 60% of the blocks contained 2 targets. A pseudorandomized stimulus sequence was generated for each subject. The order of task conditions was counterbalanced within and across subjects.

Experimental setup

Visual and auditory stimuli were presented using Cogent (John Romaya, Vision Lab, UCL; http://www.vislab.ucl.ac.uk/), running under Matlab 7.0 (MathWorks Inc.) on a Windows PC. Visual stimuli were back-projected onto a Plexiglas screen using a LCD projector (JVC Ltd., Yokohama, Japan) visible to the subject through a mirror mounted on the MR head coil. Auditory stimuli were presented using MR-compatible headphones (MR Confon GmbH). Subjects performed a behavioral task using a MR-compatible custom-built button device connected to the stimulus computer.

Behavioral measurements

Subjects' performance measures (% correct, median reaction times) were entered into repeated measurement ANOVAs or paired *t* tests. Signal sensitivity measures d' [$= Z(P_{\text{hits}}) - Z(P_{\text{false alarms}})$] were computed for the unimodal (i.e., V, Vn, A, An) and bimodal object trials (i.e., AV) during the explicit categorization task. Further, the subject-specific multisensory perceptual benefit was calculated for both designs, e.g., $d'(\text{AV}) - \max[d'(\text{V}) \text{ or } d'(\text{A})]$, and later used to predict the fMRI signals (see Data analysis). To investigate whether subjects efficiently integrated audiovisual information, we compared the empirical d' for the bimodal trials (AV) to the d' predicted by a probability summation model based on the two unimodal object conditions. The prediction of the probability summation model is calculated from the two relevant unimodal d' values under the assumption that visual and auditory information are processed independently and combined for the final behavioral decision using an "either-or" rule (Wickens, 2002). In a signal detection task, a "yes"–response is elicited when a signal is detected either in the visual or auditory modality. Thus, the decision bound of the probability summation model is formed from two lines (i.e., the two unimodal decision bounds) at right angle. We have applied this model to our two alternative forced choice categorization task by arbitrarily treating one category (e.g., tools) as signal and the other one as noise (e.g., musical instruments). The predicted hits and false alarms (for the d' of probability summation model) were computed from the unimodal conditions ([UI]: V and A; [UO]: Vn and An) as follows: Probability of hits, $P_{\text{hits}}(\text{AV}) = P_{\text{hits}}(\text{A}) + P_{\text{hits}}(\text{V}) - P_{\text{hits}}(\text{A}) \times P_{\text{hits}}(\text{V})$; probability of false alarms, $P_{\text{false alarms}}(\text{AV}) = P_{\text{false alarms}}(\text{A}) + P_{\text{false alarms}}(\text{V}) - P_{\text{false alarms}}(\text{A}) \times P_{\text{false alarms}}(\text{V})$.

An empirical $d'(\text{AV})$ that is significantly greater than predicted by the probability summation (PSM) suggests that subjects have not independently processed but integrated the information from the two input modalities to some extent (Treisman, 1998).

MRI

A 3T SIEMENS MAGNETOM TrioTim System (Siemens) was used to acquire both, T1-weighted anatomical images (176 sagittal slices, TR = 1900 ms, TE = 2.26 ms, TI = 900 ms, flip angle = 9°, FOV = 256 mm × 224 mm, image matrix = 256 × 224, voxel size = 1 mm × 1 mm × 1 mm) and T2*-weighted axial echoplanar images with blood oxygenation level-dependent (BOLD) contrast (GE-EPI, Cartesian *k*-space sampling, TR = 3080 ms, TE = 40 ms, flip angle = 90°, FOV = 192 mm × 192 mm, image matrix 64 × 64, 38 slices acquired sequentially in ascending direction, 3.0 mm × 3.0 mm × 2.6 mm voxels, interslice gap 0.4 mm). There were four sessions with a total of 245 volume images per session. The first 3 volumes were discarded to allow for T1-equilibration effects. The high-resolution anatomical image volume was acquired at the end of the experiment.

Data analysis

The data were analyzed with statistical parametric mapping (using SPM5 software from the Wellcome Department of Imaging Neuroscience, London; http://www.fil.ion.ucl.ac.uk/spm) (Friston et al., 1995). Scans from each subject were realigned using the first as a reference, unwarped, spatially normalized into MNI standard space (Evans et al., 1992), resampled to $3 \times 3 \times 3$ mm^3 voxels and spatially smoothed with a Gaussian kernel of 8 mm FWHM. The time series of all voxels were highpass filtered to 1/128 Hz. The fMRI experiment was modeled in an event related manner with regressors entered into the design matrix after convolving each event-related unit impulse with a canonical hemodynamic response function and its first temporal derivative. In addition to modeling the 14 conditions in our experiment (i.e., 4 + 3 conditions in the [UI] and [UO] 2 × 2 factorial designs with the AV condition being common to both, under each of the two task-contexts), the statistical model included instructions and the three individual target types (i.e., visual, auditory and audiovisual targets). Realignment parameters were included as nuisance covariates to account for residual motion artifacts. Condition-specific effects for each subject were estimated according to the general linear model and passed to a second-level analysis as contrasts to allow a random-effects analysis with inferences at the population level (Friston et al., 1999). This involved creating the following contrast images at the first level (averaged over sessions): (1) Superadditive audiovisual interactions separately for the [UI] and [UO] designs and separately for each task context; (2) Subadditive audiovisual interactions separately for the [UI] and [UO] designs and separately for each task context; (3) Increased superadditive effects for [UI] > [UO] design, separately for each task context; (4) Increased subadditive effects for EXPLICIT > IMPLICIT task, separately for the [UI] and [UO] design; (5) All stimuli > fixation blocks (note: the design included fixation null-events (i.e., Fix) within the stimulus blocks and fixation "baseline" blocks).

Contrast images were entered into second level ANOVAs to enable conjunction analyses across (1) the explicit and implicit tasks or (2) the [UI] and [UO] designs. In each case, we performed a conjunction (conjunction null) analysis that tested for a logical "AND" (Friston et al., 2005; Nichols et al., 2005). In addition, the superadditive interaction contrast images for the [UO] and [UI] designs were separately entered into regression analyses that used the subject's multisensory behavioral benefit, that is, the corresponding (i.e., [UO] or [UI] design) increase in perceptual (d') sensitivity, as predictors (see Noppeney et al., 2008; Holmes et al., 2008 for fMRI studies that used multisensory congruency effects to predict BOLD-responses).

There is currently still some debate about how to identify multisensory integration sites using fMRI. Several criteria such as the max criterion, the mean criterion or conjunction analyses have been proposed as methodological approaches (Calvert, 2001; Beauchamp, 2005; Goebel and van Atteveldt, 2009). However, given the limited spatial resolution of the BOLD-response none of these approaches enables the dissociation of true multisensory integration from regional convergence, i.e., where the bisensory response is equal to the sum of the two unisensory responses. Given this fundamental problem of independent unisensory neuronal populations within a particular region (i.e., voxel), we have used a more stringent methodological approach that poses response additivity as the null hypothesis and identifies multisensory integration through response nonlinearities, i.e., the interaction between visual and auditory inputs (Calvert, 2001; Calvert et al., 2001). Yet, one drawback of this interaction approach is that neurophysiological studies have also demonstrated additive combinations of inputs from multiple sensory modalities (Laurienti et al., 2005; Stanford et al., 2005). Therefore we also used regression analyses that used the subjects' multisensory behavioral benefit to predict their audiovisual interaction profiles. As we will see later, this approach allows us to identify brain regions with additive response combinations across subjects that would have otherwise evaded our interaction analysis (for further methodological discussion, see Noppeney, 2010).

To dissociate audiovisual interactions at multiple processing stages in object categorization, we used the following analysis rationale (see also supplemental Table S4, available at www.jneurosci.org as supplemental material).

Audiovisual interactions that differ for the [UI] and [UO] design, but are common to both tasks. Audiovisual integration processes that depend of the type of information that is integrated, but do not depend on the particular task-context were identified as follows:

First, we tested for superadditive (or subadditive) audiovisual interactions in the [UI] design (AV+Fix) ≠ (A+V) that were common to both explicit and implicit tasks (i.e., a conjunction analysis over task contexts). These audiovisual interactions encompass both, low-level integration processes attributable to costimulation (e.g., salience effects) and higher-order integration of object features.

Second, we identified superadditive (or subadditive) audiovisual interactions in the [UO] design (AV+N) ≠ (An+Vn) that were common to both explicit and implicit tasks (i.e., a conjunction analysis over task contexts). As the Unimodal Object information [UO] design provided low-level auditory and visual inputs in all conditions and only manipulated their informativeness by adding different amounts of noise, it controls for effects of audiovisual costimulation and selectively focuses on integration of higher-order object information.

Third, to identify areas associated with "automatic" salience effects due to audiovisual costimulation, we directly compared the audiovisual interactions of the [UI] and [UO] designs (again as a conjunction over task contexts). More specifically, we tested for superadditive interactions that were enhanced for the [UI] relative to the [UO] design (note: we subtracted audiovisual [UO] interactions that pertain only to integration of object information from [UI] interactions that emerge due to audiovisual costimulation and/or object information; the difference should therefore selectively reveal low-level interactions due to costimulation) (see supplemental Table S4, available at www.jneurosci.org as supplemental material, for analysis rationale).

Audiovisual interaction profiles predicted by subjects' multisensory behavioral benefits. To identify regions associated with the integration of higher-order features into perceptual object representations, we regressed subjects' superadditive BOLD-response interactions in the [UI] and [UO] design (explicit categorization task) on their multisensory behavioral benefit as measured by an increase in perceptual sensitivity (d').

Audiovisual interactions that depend on the task, but are common to both, the [UI] and [UO] designs. Audiovisual integration processes that depend on the task, but are common to the [UI] and [UO] designs were identified as follows.

First, we tested for superadditive (or subadditive) audiovisual interactions in the explicit categorization task that were common to both the [UI] and [UO] designs (i.e., a conjunction analysis over [UI] and [UO] designs). Please note that the interaction term (AV+Fix) ≠ (A+V) in the [UI] design is not balanced during the categorization task, since no task can be performed on the "Fix" stimuli. However, the interaction term (AV+N) ≠ (An+Vn) in the [UO] design is balanced to a high degree because of a replacement of Fix with uninformative noise trials (N). One may argue that categorization of noise stimuli that do not provide useful category information are not equivalent to categorization of degraded object stimuli in terms of the underlying cognitive processes. However, experiments have provided evidence that similar cognitive processes are involved when subjects categorize or discriminate white noise or uninformative signals. For instance, when subjects are presented with white noise stimuli, reverse correlation techniques based on their perceptual decisions were able to reveal subjects' internal object representations (Gosselin and Schyns, 2003). These results suggest that even in the absence of bottom up object information subjects perform categorization on internal object representations and treated the uninformative noise trials (N) as extremely degraded stimuli (note: task-induced processing may have even been enhanced for the N stimuli as indexed by longer response times). Hence, a conjunction analysis across [UO] and [UI] designs that forms a logical "AND" operation should therefore identify those interactions that are truly attributable to multisensory facilitation of semantic categorization.

Second, we tested for superadditive (or subadditive) audiovisual interactions in the implicit target detection task that were common to both the [UI] and [UO] designs (i.e., a conjunction analysis over [UI] and [UO] designs). During the implicit target detection task, subjects responded to rare simple target events, but not to the audiovisual object stimuli that entered into this interaction contrast. Hence, audiovisual interactions during the implicit target detection task should be influenced only to a very limited degree by differences in task-induced processing or demands on attentional resources.

Third, to identify areas associated with multisensory facilitation of semantic retrieval and categorization, we directly compared the audiovisual interactions during the explicit and implicit tasks (again as a conjunction over [UI] and [UO] designs). More specifically, we tested for subadditive interactions that were enhanced for the explicit categorization relative to the implicit processing task (as a conjunction over [UI] and [UO] designs) (see supplemental Table S4, available at www.jneurosci.org as supplemental material, for analysis rationale).

Search volume constraints

The search space (i.e., volume of interest) was constrained using orthogonal contrasts and a priori anatomical regions based on previous functional imaging findings. Each effect was tested for in two nested search volumes. The first search volume was limited to all voxels that were activated > fixation (i.e., baseline) at a threshold of $p < 0.05$, uncorrected (24,481 voxels). The second search volume (STS) was limited to the subset of activated voxels that were located within Heschl's gyri and the middle/superior temporal gyri bilaterally (2958 voxels), as defined by the AAL library (Tzourio-Mazoyer et al., 2002) using the MarsBaR (http://marsbar.sourceforge.net/) toolbox (Brett et al., 2002). Unless otherwise stated, we report activations at $p < 0.05$, corrected at the cluster level for multiple comparisons within the search volume using an auxiliary uncorrected voxel threshold of $p < 0.001$ (Friston et al., 1994). Cytoarchitectonically defined regions were assigned using the SPM-anatomy-toolbox that provides probabilistic cytoarchitectonic maps of the human brain (Eickhoff et al., 2005).

Effective connectivity analysis: dynamic causal modeling

Dynamic causal modeling (DCM) treats the brain as a dynamic input-state-output system. The inputs correspond to conventional stimulus functions encoding experimental manipulations. The state variables are neuronal activities and the outputs are the regional hemodynamic responses measured with fMRI. The idea is to model changes in the states, which cannot be observed directly, using the known inputs and outputs. Critically, changes in the states of one region depend on the states (i.e., activity) of others. This dependency is parameterized by effective connectivity. There are three types of parameters in a DCM: (1) input parameters which describe how much brain regions respond to experimental stimuli, (2) intrinsic parameters that characterize effective connectivity among regions and (3) modulatory parameters that characterize changes in effective connectivity caused by experimental manipulation. This third set of parameters, the modulatory effects, allows us to explain fMRI effects of audiovisual costimulation by changes in coupling among brain areas. Importantly, this coupling (effective connectivity) is expressed at the level of neuronal states. DCM employs a forward model, relating neuronal activity to fMRI data that can be inverted during the model fitting process. Put simply, the forward model is used to predict outputs using the inputs. The parameters are adjusted (using gradient descent) so that the predicted and observed outputs match. This adjustment corresponds to the model-fitting.

For each subject, 9 DCMs (Friston et al., 2003) were constructed. Each DCM included three regions. (1) The right Heschl's gyrus (HG; $x = 45$, $y = -21$, $z = 12$) that showed superadditive BOLD-response enhancements in the [UI] design was used as the auditory input region. (2) Since the right superior temporal sulcus is currently considered as the main audiovisual convergence area that provides feedback modulation to HG (Kayser and Logothetis, 2009; Musacchia and Schroeder, 2009) and has also been identified as a key player in previous fMRI connectivity analyses (Noesselt et al., 2007; van Atteveldt et al., 2009), we included the STS as a higher-order multisensory area in our DCM. To focus on a region within STS that is involved in multisensory processing in our paradigm, the STS region was selected based on the regression analysis showing superadditive interactions that increased with multisensory behavioral benefits across subjects (STS; $x = 51$, $y = -27$, $z = -3$). (3) The calcarine region was selected based on all stimuli > baseline [right calcarine sulcus (CaS); $x = 9$, $y = -96$, $z = -3$) as the input region for visual input (note: selecting a region according to a contrast of interest does not bias the DCM analysis). The three regions were bidirectionally connected with visual stimuli entering as extrinsic inputs to CaS and auditory stimuli to HG. Holding the intrinsic and extrinsic connectivity structure constant, the 9 DCMs manipulated the connection(s) that was (were) modulated by audiovisual costimulation in a 3 × 3 factorial manner with factors: (1) Pathway: Audiovisual costimulation modulated (a) any of the "direct" connections between CaS and HG, (b) the "indirect" connections via STS or (c) both, "direct+indirect" connections between CaS and HG. (2) Directionality: Audiovisual costimulation modulated connections (a) unidirectionally from HG, (b) unidirectionally to HG or (c) bidirectionally from and to HG (see Fig. 7*A*). In particular, these DCMs enable us to arbitrate between two main hypotheses currently advanced in multisensory research: (1) Multisensory integration in low-level areas may be mediated via recurrent loops from higher-order convergence areas such as STS as recently proposed by other authors (Noesselt et al., 2007; van Atteveldt et al., 2009). (2) The integration may be mediated via lateral connectivity between sensory areas (primary or higher-order auditory or visual areas). Since we did not detect any multisensory effects in the thalamus (most likely because of methodological limitations of fMRI), this structure was not included in the analysis. However, we note that audiovisual interactions at the thalamic level may be an additional mechanism that needs to be further investigated using fMRI acquisition methods optimized for imaging subcortical structures.

The regions were selected using the maxima of the relevant contrasts (or behavioral regression), i.e., of the random-effects analysis. Region-specific time-series (concatenated over the four sessions, adjusted for the [UO] design and confounds) comprised the first eigenvariate of all voxels within a 6 mm radius centered on these maxima. The adjustment of the region-specific time-series to the [UI] design enabled us to limit our DCM analysis to the [UI] design only (since only signal of the [UI] design was included in the first eigenvariate). In the DCM models, the timings of the onsets of the [UI] design were individually adjusted for each region to match the specific slice acquisition time (note: The adjustment of the trial onsets is different from traditional slice timing in data preprocessing that involves interpolation of the fMRI data).

Bayesian model comparison

To determine the most likely of the 9 DCMs given the observed data from all subjects, we implemented a fixed and a random effects group analysis. The fixed effects group analysis reports the posterior probability of each model and each model family. The model posterior probability was obtained by taking the product of the model evidences of each subject-specific DCM and its model prior (Penny et al., 2004). The posterior probability of a model family is the sum of the posterior probabilities of all models within a family. The model evidence as approximated by the free energy does not only depend on model fit but also model complexity. Because the fixed effects group analysis can be distorted by outlier subjects, Bayesian Model Selection was also implemented in a random effects group analysis using a hierarchical Bayesian model that estimates the parameters of a Dirichlet distribution over the probabilities of all models considered (implemented in SPM8). These probabilities define a multinomial distribution over model space enabling the computation of the posterior probability of each model given the data of all subjects and the models considered. To characterize our Bayesian Model Selection results at the random effects level, we report the exceedance probability of one model being more likely than any other model tested (Stephan et al., 2009). The exceedance probability quantifies our belief about the posterior probability that is itself a random variable. Thus, in contrast to the expected posterior probability, the exceedance probability also depends on the confidence in the posterior probability. For the optimal model, the subject-specific intrinsic, modulatory and extrinsic effects were also entered into one-sample *t* tests. This allowed us to summarize the consistent findings from the subject-specific DCMs using classical statistics.

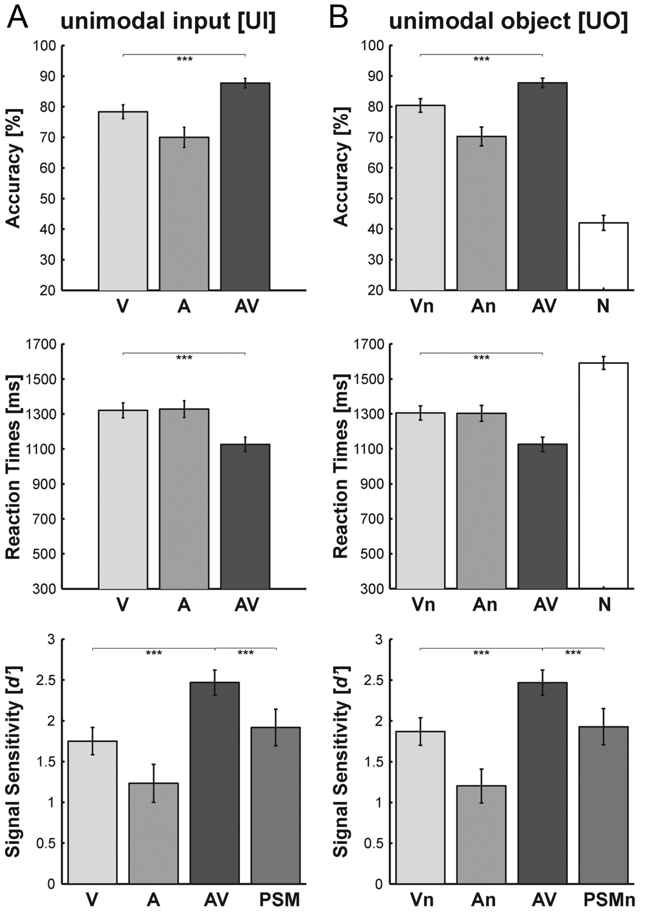

Figure 2. Bar plots showing categorization performance. ***A***, Unimodal Input design, i.e., unimodal (V, A) and bimodal (AV) stimuli. ***B***, Unimodal Object information design, i.e., unimodal (Vn, An), bimodal (AV) and audiovisual noise (N) stimuli. Top: Accuracy rates (across subject mean ± SEM). Middle: Reaction times (across subject mean ± SEM). Bottom: Signal sensitivity *d'* (across subject mean ± SEM). The PSM/PSMn bars represent the prediction of probability summation models (see Materials and Methods). ***Significant at $p < 0.001$.

Model comparison and statistical analysis of connectivity parameters of the optimal model enabled us to address the following three questions: First, comparing the three pathway model families, i.e., (1) direct, (2) indirect, and (3) indirect+direct, we asked whether the superadditive responses in HG are mediated by dynamic changes in (1) direct connectivity between visual and auditory areas, (2) indirect connectivity via STS as a higher-order multisensory convergence region or (3) both mechanisms. Please note that direct connectivity refers to functional/effective connectivity alone; it does not imply that this direct functional connectivity is mediated by direct anatomical connectivity between V1 and A1 (it can also be mediated by higher-order visual areas and planum temporale). Second, comparing the three "directionality" model families, we asked whether the superadditive responses in HG are mediated by changes in connectivity (1) "from," (2) "to," or (3) "from and to" HG. Third, we determined the most likely of the 9 DCMs given the observed data from all subjects at the fixed and random effects group level.

Table 1. Audiovisual interactions that differ for the unimodal input and unimodal object information design, but are common to both tasks

	MNI coordinates					
Regions	*x*	*y*	*z*	*z*-score (peak)	*p*-value$_c$, (cluster)	Number of voxels
Superadditive audiovisual interactions: [UI] design						
Conjunction across tasks: $[(AV_{EXPL} + Fix_{EXPL}) - (V_{EXPL} + A_{EXPL})] \cap [(AV_{IMPL} + Fix_{IMPL}) - (V_{IMPL} + A_{IMPL})]$						
R. Heschl's gyrus	45	−21	12	4.40	0.027	53
	45	−24	18	4.39		
Superadditive audiovisual interactions: [UO] design						
Conjunction across tasks: $[(AV_{EXPL} + N_{EXPL}) - (Vn_{EXPL} + An_{EXPL})] \cap [(AV_{IMPL} + N_{IMPL}) - (Vn_{IMPL} + An_{IMPL})]$						
No significant effects						
Superadditive audiovisual interactions: [UI] > [UO]						
Conjunction across tasks: EXPLICIT ∩ IMPLICIT						
R. Heschl's gyrus	45	−24	15	3.18	0.001 #	

p-value$_c$, Corrected at cluster level for multiple comparisons within the search volume of all voxels activated at $p < 0.05$, uncorrected (24,481 voxels). Auxiliary uncorrected voxel threshold of $p < 0.001$. #Uncorrected *p*-value at peak voxel. R., Right.

Results

Behavioral data

Explicit task

Subjects categorized visual (V, Vn), auditory (A, An) and audiovisual (AV) object stimuli as well as the audiovisual noise condition (N) as tools or musical instruments (Fig. 2). Multisensory behavioral benefits were assessed by comparing performance accuracy and reaction times of the bimodal condition to the best (i.e., most accurate and fastest) unimodal condition (paired *t*-tests). Both for the [UO] and the [UI] design, we observed significant increases in performance accuracy (acc) and decreases in reaction times (rt) for the bimodal relative to the best unimodal (i.e., visual) condition ($[UI]_{acc}$: $t_{20} = 6.2$; $p < 0.001$; $[UO]_{acc}$: $t_{20} = 5.5$; $p < 0.001$; $[UI]_{rt}$: $t_{20} = -9.9$; $p < 0.001$; $[UO]_{rt}$: $t_{20} = -7.9$; $p < 0.001$). Similarly, subjects' perceptual sensitivity (d') was significantly increased for the bimodal relative to the best unimodal condition ([UI]: $t_{20} = 6.1$; $p < 0.001$; [UO]: $t_{20} = 5.3$; $p < 0.001$). Importantly, the d' of the bimodal condition was significantly larger than the d' predicted by a probability summation model (PSM) of the two unimodal conditions ([UI]: $t_{20} = 4.1$; $p < 0.001$; [UO]: $t_{20} = 3.9$; $p < 0.001$). The prediction of the PSM is derived from the two unimodal d' sensitivity measures under the assumption that on each trial visual and auditory signals are processed independently and combined in the behavioral decision using an "either-or-rule". Performance accuracy for the bimodal trials was better than predicted by the probability summation model, suggesting that subjects efficiently integrated visual and auditory information during object categorization (Fig. 2).

Implicit task

During the implicit task, subjects passively attended to the task-irrelevant object stimuli (i.e., tools and musical instruments), while responding to simple detection items i.e., a visual circle, an auditory tone or an audiovisual circle+tone. Subjects achieved ceiling performance (mean ± SEM) for the target items with detection accuracies (visual: 99.0 ± 0.6%; auditory: 98.3 ± 0.7%; audiovisual: 99.8 ± 0.2%) and reaction times (visual: 449 ± 12 ms; auditory: 431 ± 16 ms; audiovisual: 391 ± 14 ms). A one-way repeated measurement ANOVA of the target conditions (auditory, visual, audiovisual) identified a significant main effect in terms of accuracy ($F_{(1.8,36.7)} = 3.6$; $p < 0.05$) and reaction times ($F_{(1.3,26.1)} = 46.8$; $p < 0.001$) after Greenhouse-Geisser correction. *Post hoc* comparisons (Bonferroni corrected) revealed no significant increase in multisensory detection performance (with respect to the best unimodal condition), but multisensory response facilitation in terms of reaction times, i.e., audiovisual responses were faster compared with auditory and visual ones.

Neuroimaging data

As described in more detail in the methods section, the analysis was performed in three steps. First, we tested for superadditive and subadditive audiovisual interactions that were common to both, explicit and implicit tasks (i.e., a conjunction analysis over task contexts) separately for the [UI] and [UO] design. Second, we performed a second level regression analysis in which we used subjects' multisensory behavioral benefit in object categorization to predict superadditive interactions. Third, we tested for subadditive and superadditive audiovisual interactions that were common to both, the [UI] and [UO] design (i.e., a conjunction analysis over designs) separately for the explicit categorization and implicit processing tasks. We further characterized the response pattern of each region according to (1) the type of interaction (i.e., superadditive vs subadditive), (2) the magnitude of the bimodal response relative to the unimodal responses (i.e., multisensory enhancement vs suppression), (3) the sensory dominance (visual, auditory, none) and (4) task dependence.

Audiovisual interactions that differ for the [UI] and [UO] design, but are common to both tasks

Automatic audiovisual interactions were identified by testing for superadditive (and subadditive) audiovisual interactions that were common to both, explicit and implicit tasks (i.e., a logical "AND" conjunction over tasks) separately for the [UI] and [UO] designs (Table 1).

For the [UI] design, superadditive interactions were found in right HG extending into the posterior insula and parietal operculum. Based on probabilistic cytoarchitectonic maps, parts of the activations were localized in subdivisions (Te1.1, Te1.0, Te1.2) of human primary auditory cortex (Morosan et al., 2001). More specifically, 22.8% of the cluster in Figure 3*A* activated 26.3% of the total Te1.1. Here, the sum of the bimodal and fixation conditions (i.e., AV+Fix) significantly exceeded the sum of the unimodal (i.e., V+A) BOLD-responses regardless of task contexts (Fig. 3*A*; Table 1). Interestingly, while visual input alone leads to deactivation in auditory areas, in the context of auditory stimulation, it amplifies the auditory response. This audiovisual enhancement relative to the maximal unimodal, i.e., auditory, BOLD-response was common to both tasks (Fig. 3*B*, top). Additional *post hoc* analyses further dissociated two mechanisms underlying the audiovisual response enhancement in HG by comparing (1) An to A and (2) AV to An (as mean across tasks). (1) An (1.59 ± 0.23) was significantly enhanced relative to A (1.24 ± 0.21) indicating that the response amplification in auditory cortex does not depend on object information, but can also be found for auditory input with concurrent visual noise (An > A: $t_{20} = 2.0$; $p < 0.05$; 1-tailed). (2) However, we also found a trend toward enhancement for AV (1.77 ± 0.22) relative to An (AV > An: $t_{20} = 1.5$; $p = 0.078$; 1-tailed). This enhancement may originate from additional audiovisual temporal structure attributable to the action sequences embedded in the audiovisual noise streams. The audiovisual action sequences may provide additional synchrony cues important for low-level audiovisual integration. Note that the onset cues of the noise streams are truly synchronous. In contrast, the transient changes of the action sequences embedded in the audiovisual noise streams are often characterized by an auditory lag according to the natural statistics of action sequences (e.g., the visual motion of a hammer usually precedes the sound of its strike).

In contrast to the superadditive responses of the [UI] design, no superadditive interactions were observed for the [UO] design, that controlled for salience enhancing effects of costimulation and focused selectively on integration of higher-order object information (Fig. 3*B* bottom). Indeed, when directly compared, the superadditive interactions in the right HG were significantly greater for the [UI] than the [UO] design again regardless of task contexts at an uncorrected level of significance (Table 1).

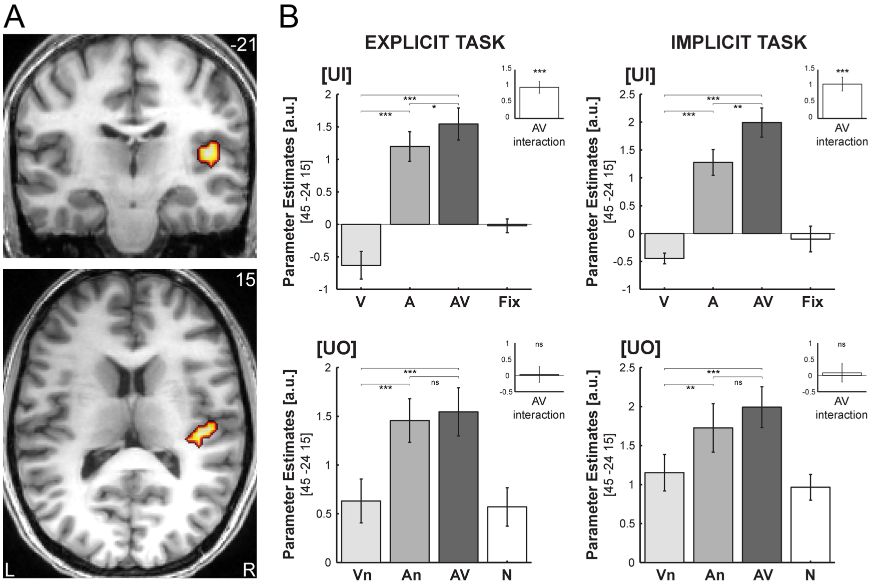

Figure 3. ***A***, Superadditive interactions common to both tasks in right HG for the [UI] design on coronal and transversal slices of a subjects' normalized structural image. Height threshold: $p < 0.001$, uncorrected for illustration purposes and inclusively masked with all stimuli > baseline at $p < 0.05$, uncorrected. ***B***, Parameter estimates (across subjects' mean ± SEM) for (1) the unimodal (V, A), bimodal (AV), and fixation (*Fix*) conditions of the [UI] design (top) and (2) the unimodal (Vn, An), bimodal (AV), and audiovisual noise (N) conditions of the [UO] design (bottom) in the explicit (left) and implicit (right) tasks at a given coordinate location. Insets show the parameter estimates for the audiovisual interactions in each design and task. The bar graphs represent the size of the effect in nondimensional units (corresponding to percentage whole brain mean). Asterisks indicate significance at $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) uncorrected.

Collectively, this response pattern suggests that superadditive audiovisual interactions in low-level auditory areas mediate (stimulus driven) automatic salience effects that emerge due to costimulation of spatiotemporally aligned auditory and visual inputs.

No significant subadditive interactions were found for the [UI] or the [UO] design as a conjunction across task contexts.

Audiovisual interaction profiles predicted by subjects' multisensory behavioral benefits

To identify integration processes of higher-order object information, we used subjects' multisensory behavioral benefits (d') (i.e., increased categorization performance for the AV relative to the best unimodal conditions) to predict their superadditive BOLD-responses separately for the [UI] and [UO] designs (limited to the categorization task).

In the [UI] design, subjects' audiovisual benefit significantly predicted superadditive activations in the right superior temporal sulcus (STS), right planum temporale (PT) and left posterior intraparietal sulcus (IPS). In other words, STS/PT and IPS showed increasingly superadditive BOLD-responses with increasing audiovisual benefits: subjects with high audiovisual benefits showed predominantly positive (i.e., superadditive) audiovisual interactions, whereas those with less or no audiovisual benefits showed negative (i.e., subadditive) interactions (Fig. 4; Table 2). Averaged across subjects, both, STS/PT and IPS exhibited primarily additive response combinations with a trend toward subadditivity (see Fig. 6*B*). Hence, only the additional regression analysis based on intersubject variability enabled us to reveal audiovisual integration in those region, because additive response combinations predominated when averaging across subjects. In STS/PT and IPS,

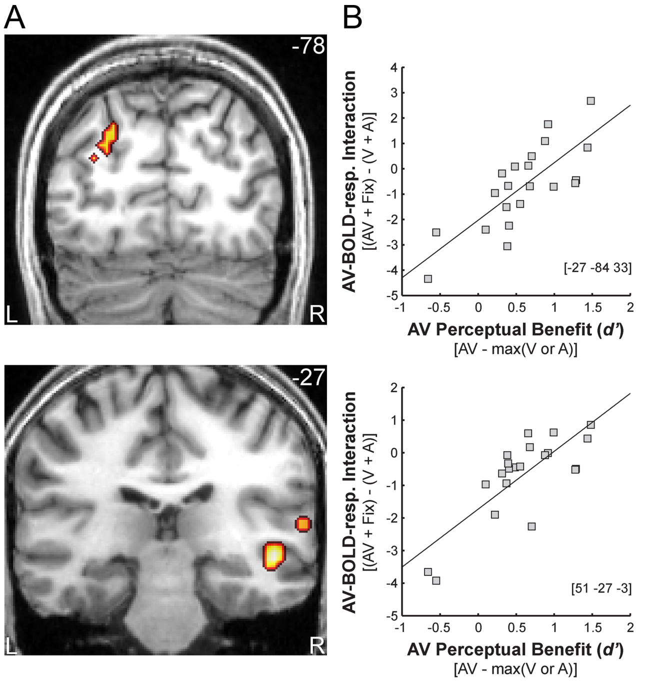

Figure 4. ***A***, Audiovisual BOLD-response interactions that were predicted by subjects' multisensory behavioral benefit in left posterior IPS (top), right STS and PT (bottom) for the [UI] design in the explicit categorization task on coronal slices of a subjects' normalized structural image. Height threshold: $p < 0.001$, uncorrected for illustrational purposes. Extent threshold >15 voxels. ***B***, Scatter plot depicting the regression of subjects' superadditivity (ordinate) on the perceptual benefit (abscissa) across subjects in IPS (top) and STS for given coordinate locations (bottom) (see Materials and Methods for further details).

Table 2. Audiovisual interactions profiles predicted by subjects' multisensory behavioral benefits

	MNI coordinates					Number
Regions	x	y	z	z-score (peak)	p-value$_c$, (cluster)	of voxels
Regression analysis: AV-BOLD-response interaction contrast versus multisensory perceptual benefit						
[UI] design EXPLICIT task: $[(AV_{EXPL} + Fix_{EXPL}) - (V_{EXPL} + A_{EXPL})]$ versus d'(AV) − best[d'(A or V)]						
L. intraparietal sulcus (posterior)	−27	−84	33	4.10	0.016	46
R. superior temporal sulcus (middle)	51	−27	−3	4.18	0.020[#]	22
R. planum temporale	63	−33	12	3.84	0.032[#]	18

p-value$_c$, Corrected at cluster level for multiple comparisons within the search volume of all voxels activated at $p < 0.05$, uncorrected (24,481 voxels); [#] the STS search volume (2958 voxels). Auxiliary uncorrected voxel threshold of $p < 0.001$. L., Left; R., right.

Table 3. Audiovisual interactions that depend on the task but are common to both the unimodal input and unimodal object information designs

	MNI coordinates					Number
Regions	x	y	z	z-score, (peak)	p-value$_c$, (cluster)	of voxels
Subadditive audiovisual interactions: EXPLICIT task						
Conjunction across [UI] and [UO] designs: $[(V + A) - (AV + Fix)] \cap [(Vn + An) - (AV + N)]$						
L. inferior precentral sulcus	−45	9	27	6.53	0.000	871
L. inferior frontal sulcus	−39	33	18	5.38		
L. ventral precentral gyrus	−45	3	33	4.89		
L. insula (anterior)	−30	30	6	4.85		
L. caudate nucleus (head)	−15	−3	15	4.60		
L. thalamus (ventral lateral posterior)	−12	−18	9	4.27		
L. putamen	−21	6	−3	4.20		
L. inferior frontal gyrus	−48	9	12	4.08		
L. inferior temporal/fusiform gyrus	−45	−54	−15	6.04	0.000	191
L. middle temporal gyrus/superior temporal sulcus (posterior)	−51	−57	6	3.39		
L. intraparietal sulcus	−27	−63	45	4.53	0.000	169
L. intraparietal sulcus (anterior)	−30	−42	39	3.84		
L. intraparietal sulcus (posterior)	−27	−78	36	3.27		
R. caudate nucleus (head)	12	3	3	5.44	0.001	124
R. putamen	27	12	−3	3.48		
R. thalamus (ventral lateral posterior)	9	−12	9	3.16		
R. inferior frontal gyrus	48	27	21	4.76	0.000	227
R. insula (anterior)	33	27	0	4.58		
R. ventral precentral gyrus	45	3	33	3.39		
Subadditive audiovisual interactions: IMPLICIT task						
Conjunction across [UI] and [UO] designs: $[(Vn + An) - (AV + N)] \cap [(V + A) - (AV + Fix)]$						
No significant effects						
Subadditive audiovisual interactions: EXPLICIT task > IMPLICIT task						
Conjunction across designs: $[UI] \cap [UO]$						
L. inferior precentral sulcus	−45	9	21	5.10	0.002	101
L. inferior frontal sulcus	−45	33	12	4.38	0.002	99
L. inferior frontal sulcus	−39	33	15	4.30		

p-value$_c$, Corrected at cluster level for multiple comparisons within the search volume of all voxels activated at $p < 0.05$, uncorrected (24,481 voxels). Auxiliary uncorrected voxel threshold of $p < 0.001$. L., Left; R, right.

audiovisual BOLD-responses were suppressed relative to the most effective unimodal, but enhanced relative to the least effective one. Yet, the two regions differed in terms of their sensory preference with STS/PT being auditory dominant and IPS visually dominant (see Fig. 6*B*). The relationship between multisensory behavioral benefits and superadditive BOLD-responses in STS/PT and IPS suggests that this network of sensory dominant regions may integrate audiovisual object information into perceptual representations that are relevant for categorization.

For the [UO] design, the regression analysis did not reveal any significant activation after correcting for multiple comparisons. However, at an uncorrected level, we observed effects in the neural system identified in the [UI] design including the left IPS ($x = -24$, $y = -84$, $z = 39$; $z_{\mathrm{peak}} = 2.71$; $p_{\mathrm{uncorr}} = 0.003$), right STS (e.g., $x = 57$, $y = -18$, $z = -6$; $z_{\mathrm{peak}} = 1.97$; $p_{\mathrm{uncorr}} = 0.025$) and right PT (e.g., $x = 60$, $y = -21$, $z = 6$; $z_{\mathrm{peak}} = 1.70$; $p_{\mathrm{uncorr}} = 0.045$).

Audiovisual interactions that depend on the task, but are common to both, the [UI] and [UO] designs

We tested for subadditive (and superadditive) audiovisual interactions that were common to both, the [UI] and [UO] design (i.e., a conjunction analysis over [UI] and [UO]) separately for the explicit categorization and implicit processing tasks. No significant superadditive interactions were observed in either of the two task contexts. Furthermore, no subadditive interactions were observed for the implicit task, i.e., when subjects passively perceived the stimuli. However, for the explicit categorization task we observed subadditive interactions in a widespread frontal, temporal and parietal system including the inferior frontal sulci and gyri (IFS/IFG), left intraparietal sulcus (IPS), and inferior and middle temporal gyri (ITG, MTG) (Table 3, top; supplemental Fig. S1, available at www.jneurosci.org as supplemental material). In all of these areas, the sum of the unimodal BOLD-responses significantly exceeded the sum of the audiovisual and fixation (or audiovisual control noise) conditions both, for the [UI] (i.e., A+V > AV+Fix) and the [UO] designs (i.e., An+Vn > AV+N). In the IFS and the inferior precentral sulcus (iPrCS), the audiovisual BOLD-response was even suppressed relative to both, auditory and visual responses (Fig. 5*B*, left). The direct comparison between the two tasks corroborated that the subadditive interactions were indeed enhanced for the categorization task in the ventrolateral prefrontal cortex (vlPFC) (Table 3 bottom). These suppressive interactions were observed for both, the [UI] design that manipulates the presence/absence of sensory inputs and for the [UO] design that manipulates only the availability of object information. These results cannot be attributed to an unbalanced interaction term because this was present only in the [UI] design (note: the interaction term of the [UO] design was appropriately balanced by providing the subjects with noise (N) instead of fixation (Fix) trials). Furthermore, we did not only observe subadditive but suppressive interactions (i.e., audiovisual BOLD-responses were smaller than any of the two unimodal responses). This response profile cannot be attributed to an imbalance in the interaction term. Thus, the common audiovisual interactions for the [UI] and [UO] designs demonstrate the sensitivity of the vlPFC for higher-order object information during object categorization.

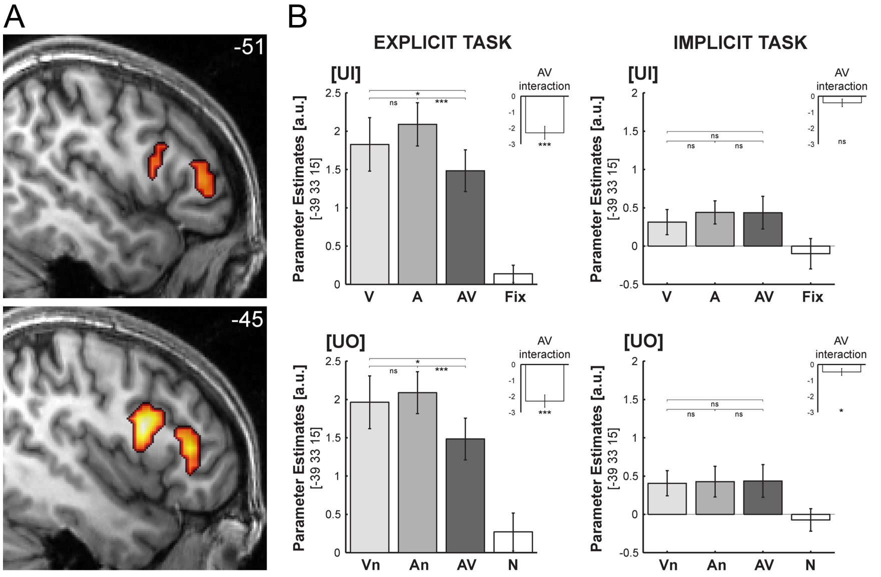

Figure 5. ***A***, Subadditive interactions (that were stronger for explicit than implicit task) common to [UI] and [UO] designs in left IFS extending into iPrCS on sagittal slices of a subject's normalized structural image. Height threshold: $p < 0.001$, uncorrected for illustrational purposes. ***B***, Parameter estimates (across subjects' mean ± SEM) of the IFS for the (1) unimodal (V, A), bimodal (AV), and fixation (Fix) conditions of the [UI] design (top) and (2) unimodal (Vn, An), bimodal (AV) and audiovisual noise (N) conditions of the [UO] design (bottom) in the explicit (left) and implicit (right) tasks at given coordinate location. Insets show the parameter estimate for the audiovisual interactions in each design and task. The bar graphs represent the size of the effect in nondimensional units (corresponding to percentage whole brain mean). Asterisks indicate significance at $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) uncorrected.

Even though no significant differences were found between the auditory and visual responses at the peak coordinates in IFS and iPrCS, additional analyses revealed sensory preferences within the vlPFC: auditory-dominant regions were located more ventrally and anterior to visual-dominant regions (Fig. 6*C*). Interestingly, the peaks of the subadditive interactions were located in the transition zone between regions with auditory and visual preferences in line with electrophysiological investigations in nonhuman primates (Romanski, 2007) or other species (Wallace et al., 2004). This suggests that the vlPFC may not only process already integrated information but also be genuinely involved in integrating auditory and visual object information that is provided from e.g., STS and IPS via dorsal and ventral streams. Collectively, these results suggest that suppressive interactions in vlPFC may reflect audiovisual facilitation of semantic retrieval and categorization that enable the selection of an appropriate response.

Summary of conventional regional SPM results

Our stimulus and task manipulations dissociated three patterns of audiovisual interactions at distinct levels of the cortical sensory processing hierarchy: (1) In HG, superadditive audiovisual interactions were observed for integrating low-level sensory inputs ([UI] design) but not higher-order object information ([UO] design). These superadditive interactions were observed both, when subjects actively categorized or passively perceived the object stimuli. Thus, superadditive auditory response amplification is of an automatic nature and may mediate low-level audiovisual salience effects (Fig. 6*A*). (2) In auditory-dominant STS/PT and visual-dominant IPS, the pattern of interaction (i.e., superadditive vs subadditive) depended on subjects' multisensory benefit in categorization performance implicating this network of regions in the perceptual formation of audiovisual object representations (Fig. 6*B*). (3) Finally, in vlPFC, suppressive audiovisual interactions were selective for the explicit categorization task and found when subjects integrate audiovisual input [UI] or audiovisual object information [UO]. The suppressed audiovisual response relative to both unimodal responses may thus reflect multisensory facilitation of semantic categorization and response selection (Fig. 6*C*).

Dynamic causal modeling

Figure 7*A* shows the 9 potential DCM candidates that differed with respect to the connection(s) that were modulated by audiovisual costimulation. Comparing pathway families of direct (M1, M2, M3) versus indirect (M4, M5, M6) versus direct+indirect (M7, M8, M9) models provided strong evidence for the direct+indirect family of models where audiovisual interactions are mediated by two distinct mechanisms: (1) direct interactions between auditory and visual areas and (2) indirect influences via recurrent loops from STS (posterior probability for direct+indirect model family > 0.94). Comparing families of different directionalities suggests that audiovisual interactions are primarily mediated by modulations of the efferent connections from HG (posterior probability for "unidirectional from HG" model family > 0.99). Both fixed and random effect analysis revealed model 8 as the optimal of the 9 models tested (FFX analysis: posterior probability for M8 > 0.94; RFX analysis: exceedance probability > 0.36) (Fig. 7*B*). Figure 7*C* shows the intrinsic and extrinsic connectivity structure and the change of connectivity strength due to audiovisual costimulation for the optimal DCM 8. Not surprisingly, auditory stimulation induces a positive activation in HG via extrinsic connectivity of auditory input to HG and visual stimulation induces a positive activation in CaS via extrinsic connectivity of visual input to HG. Audiovisual costimulation (AV) modulates the efferent direct and indirect connections from HG to STS and CaS. In particular, the STS becomes less sensitive to auditory input in the context of visual input, which in turn leads to a disinhibition of HG. Further, audiovisual costimulation reduces the inhibitory influence from auditory to visual cortices.

Discussion

The present study dissociated the functional contributions of three distinct sets of cortical regions to multisensory categorization of eco-

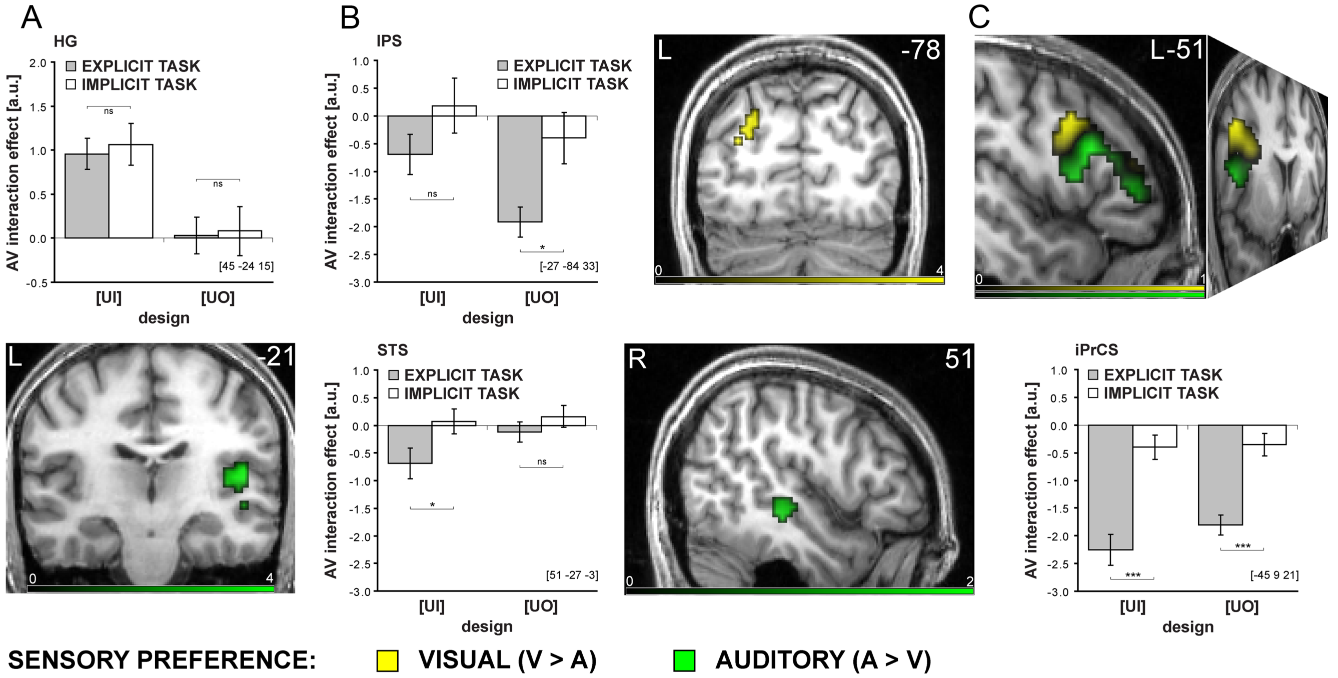

Figure 6. Cortical hierarchy of audiovisual interactions: characterization of cortical regions according to (1) sensory preference, (2) profile of audiovisual interaction, and (3) task dependence. The color codes the sensory preference, i.e., activation difference between unimodal visual vs auditory activations (green, A > V; yellow, V > A) in the [UI] design during the explicit categorization task. The bar graphs show the audiovisual interaction effects for [UI] and [UO] designs in the explicit and implicit task conditions at given coordinate locations. ***A***, Right HG is auditory dominant and shows superadditive interactions only in the [UI] design common to both task contexts. ***B***, In the left posterior IPS (top) and right STS (bottom) superadditive interactions are predicted by subjects' multisensory perceptual benefit in the [UI] design during the explicit categorization task. The STS is auditory-dominant and the IPS is visually dominant, both partly task dependent. ***C***, The vlPFC showed suppressive interactions for both the [UI] and [UO] design during the explicit categorization task (Table 3 top; supplemental Fig. S1, available at www.jneurosci.org as supplemental material). The peaks of the subadditive interactions (Table 3, bottom; Fig. 5*A*) within the iPrCS and IFS were located in the transitions zones between regions with auditory (green) and visual (yellow) preferences. Asterisks indicate significance at $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) uncorrected.

logically valid dynamic objects. First, audiovisual costimulation enhanced the stimulus salience through amplification of the neuronal responses in primary auditory cortex. Second, audiovisual interactions in higher-order association areas are related to subject's audiovisual benefit in categorization performance and may thus sustain the formation of audiovisual object representations. Third, task-dependent suppressive interactions in the ventrolateral prefrontal cortex may reflect multisensory facilitation of semantic categorization.

Automatic superadditive interactions in primary auditory cortex due to costimulation

Recent fMRI and neurophysiological research have accumulated evidence for multisensory interactions in low-level auditory areas (Foxe et al., 2002; Schroeder and Foxe, 2002; van Atteveldt et al., 2004; Ghazanfar et al., 2005; Lehmann et al., 2006; Kayser et al., 2007; Lakatos et al., 2007; Martuzzi et al., 2007). However, the functional role of these interactions is currently unclear. Here, we compared multisensory interactions for two 2 × 2 factorial designs that manipulated either (1) the presence/absence or (2) the informativeness of the sensory inputs. We demonstrate that auditory responses in Heschl's gyrus are modulated by the presence/absence of a secondary visual input, but not by its informativeness. While visual input alone leads to a deactivation in primary auditory cortex (Laurienti et al., 2002), it amplifies the response to a concurrently presented auditory input, leading to superadditive interactions. Even though our study did not manipulate the temporal or spatial relationships of visual and auditory inputs explicitly, neurophysiological studies in nonhuman primates have suggested that multisensory integration processes in primary sensory cortices are governed by tight temporal (Kayser et al., 2008) and also spatial constraints (Lakatos et al., 2007). Collectively, these results suggest that bisensory costimulation within a narrow spatiotemporal window leads to superadditive interactions that enhance stimulus salience (Kayser et al., 2005) and thus enable a coarse initial audiovisual scene-segmentation (Foxe and Schroeder, 2005). Alternatively, superadditive interactions may be caused by attentional modulation. While we cannot fully exclude any potential top-down effects (Petkov et al., 2004), e.g., due to spatial attention (Busse et al., 2005; Fairhall and Macaluso, 2009), the superadditive interactions in primary auditory cortex were observed equally for both, explicit categorization and passive exposure. These results point to automatic audiovisual integration mechanisms that are relatively immune to cognitive top-down effects. In line with our results, recent neurophysiological studies in the macaque have also demonstrated that audiovisual interactions in primary and secondary auditory regions depend neither on the particular stimulus characteristics nor the monkey's cognitive state (Kayser et al., 2008).

Multisensory interactions in low-level auditory regions can be mediated by several types of functional neural architectures including feedforward thalamocortical, direct connections between sensory areas and feedback from higher-order association areas such as IPS or STS (Schroeder et al., 2003; Driver and Noesselt, 2008). Combining Dynamic Causal Modeling and Bayesian Model Comparison, we compared models that embodied crossmodal effects via direct V1–A1 connectivity, indirect feedback connectivity from STS or via both mechanisms. Despite increased model complexity, the DCMs that allowed for both direct and indirect mechanisms underlying audiovisual integration outperformed the more constrained single mechanism models. First, audiovisual integration effects in HG were mediated by signals passing in a recurrent loop between HG and STS (see supplemental discussion, available at www.jneurosci.org as supplemental material). Evidence for indirect feedback influences from STS on PT and HG have been previously provided by fMRI studies using Granger Causality (van Atteveldt et al., 2009) and Directed Information Transfer (Noesselt et al., 2007). Second, our DCM demonstrated that audiovisual costimulation also modulated the direct connectivity between auditory and visual areas. These direct modulatory effects are in line with previous EEG and intracranial recordings in humans showing very early audiovisual integration effects (Giard and Peronnet, 1999; Molholm et al., 2004; Besle et al., 2008). Interestingly, all audiovisual integration effects are expressed in efferent connections from the auditory areas (HG and STS). This may reflect the fact that the auditory noise signal arrives first and resets the sensitivity of its target regions when auditory and visual inputs are presented concurrently. Neurophysiological recordings have suggested that this response amplification in primary visual and auditory cortices of nonhuman primates may rely on a phase resetting of ongoing neuronal oscillations with the latency difference between the component signals determining the direction of the phase resetting influence (Lakatos et al., 2007, 2008; Kayser et al., 2008).

Integration in STS, IPS, and PT depends on subjects' audiovisual benefit in categorization accuracy

Numerous neurophysiological and neuroimaging studies in human and nonhuman primates have implicated the STS and IPS in audiovisual integration (Calvert et al., 2000; Schroeder and Foxe, 2002; Macaluso et al., 2003; Wright et al., 2003; Beauchamp et al., 2004b; Barraclough et al., 2005; Miller and D'Esposito, 2005; Saito et al., 2005; Avillac et al., 2007; Ghazanfar et al., 2008; Sadaghiani et al., 2009; Stevenson and James, 2009). Indeed, their extensive bidirectional anatomical connectivity to visual and auditory areas renders them ideal for integrating inputs from multiple senses (Neal et al., 1990; Seltzer and Pandya, 1994; Falchier et al., 2002; Rockland and Ojima, 2003). In contrast to the stimulus driven superadditive interactions in primary auditory cortex, audiovisual integration in the STS, PT and IPS depended on subjects' performance in semantic categorization. While subjects that did not benefit from audiovisual integration showed subadditive interactions, subjects with improved performance during the audiovisual conditions exhibited additive and even superadditive integration profiles. The relationship between subjects' behavioral benefit and audiovisual interactions suggests that this network of STS, PT and IPS collectively integrates higher-order features into object percepts according to their auditory (STS, PT) or visual (IPS) dominance.

Furthermore, these findings demonstrate that the neural patterns of audiovisual interaction (i.e., superadditive vs subadditive) are functionally relevant and determine subjects' improvement in categorization accuracy (Werner and Noppeney, 2009). Thus, superadditive responses in higher-order association areas seem to mediate multisensory benefits at the level of object recognition,

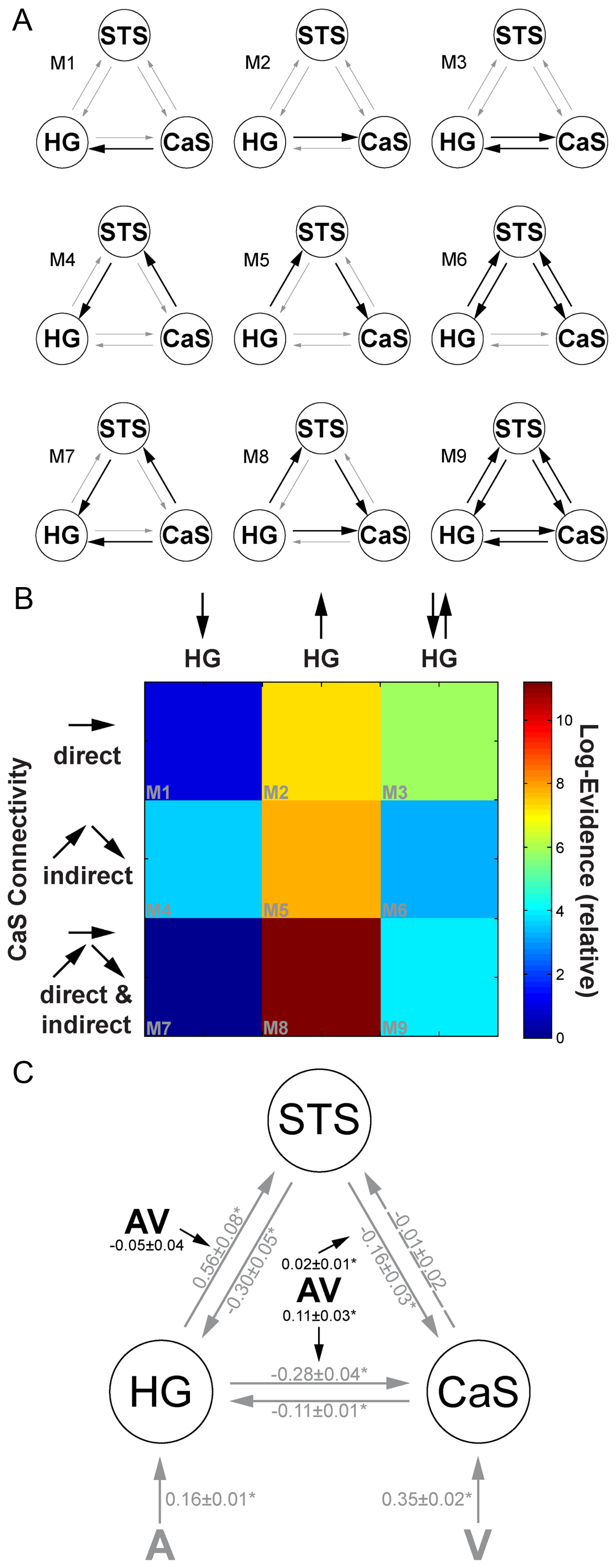

Figure 7. ***A***, The 9 candidate dynamic causal models manipulating the connection(s) that was (were) modulated by audiovisual costimulation (bold arrows). In M1 to M3, audiovisual costimulation modulated the direct connections between CaS and HG, in M4 to M6 it influenced the indirect connectivity via STS. In M7 to M9, audiovisual costimulation modulated both direct and indirect connectivity between HG and CaS. ***B***, The relative log-likelihood for each of the 9 tested DCMs that differed with respect to the connection(s) that was (were) modulated by audiovisual costimulation. ***C***, In the optimal dynamic causal model 8 [(highest posterior

←

probability (FFX analysis) and highest exceedance probability (RFX analysis)], audiovisual costimulation (AV), modulated the efferent direct and indirect connections between HG and CaS. Values are the across-subjects' mean (±SEM) of intrinsic, extrinsic and modulatory connection strength (*indexes significant at $p < 0.05$). The modulatory parameter (AV) quantifies how audiovisual costimulation changes the values of the intrinsic connections.

in a similar manner as the known improvements in orientation behavior are driven by superadditive interactions in the superior colliculus (Stein and Stanford, 2008).

Semantic categorization of audiovisual objects in ventrolateral prefrontal cortex

The neural systems underlying multisensory integration have previously been characterized primarily during passive exposure to exclude decisional and task-related processes (Calvert et al., 2000, 2001; Wright et al., 2003; van Atteveldt et al., 2004; Noesselt et al., 2007). However, multisensory perception forms the basis for our interactions with our natural multisensory environment. For instance, audiovisual integration provides an obvious survival benefit for animals that need to categorize an approaching individual as prey or foe to select an appropriate action. The present study demonstrates that multisensory categorization (but not passive exposure) induces subadditive interactions in the ventrolateral prefrontal cortex (vlPFC). Task-dependent subadditive interactions may emerge in the [UI] interaction design (i.e., presence/absence of audiovisual inputs) because of methodological problems [i.e., no task can be performed on fixation trials rendering the interaction term unbalanced and biased (Beauchamp et al., 2004a)]. However, the present study reveals subadditive interactions not only for the [UI] but also for the [UO] design that balanced the interaction term by enabling active responses to be made in all trials. Furthermore, it is important to note that in line with recent reports about suppressive interactions in the vlPFC for audiovisual monkey vocalizations (Sugihara et al., 2006), we show significant suppressions of audiovisual BOLD-responses relative to both unimodal conditions, which cannot be explained away as an artifact of the interaction analysis or by particularities of the audiovisual noise condition (see Materials and Methods). Instead, it suggests that vlPFC is involved in audiovisual integration processes and/or receives already integrated information from e.g., the STS that is known to be reciprocally connected with vlPFC (Romanski, 2004). In our study, the task-dependent suppressive interactions were located in the transition zones (Wallace et al., 2004) between auditory- and visual-dominant regions in the vlPFC (Romanski, 2007). This specific location suggests that these areas may indeed play an essential role in audiovisual integration per se rather than just processing already integrated information. The bimodal response suppression in vlPFC may reflect facilitation of semantic categorization and response selection (Vandenberghe et al., 1996; Noppeney and Price, 2002; Sabsevitz et al., 2005) similar to the well known phenomenon of repetition suppression indexing behavioral priming (Dolan et al., 1997; George et al., 1999; Henson, 2003). Audiovisual integration processes in vlPFC may thus enable us to rapidly categorize auditory and visual stimuli and select the appropriate action in our natural multisensory environment.

Conclusions

The distinct patterns of audiovisual interactions enabled us to dissociate the functional contributions of three sets of regions to audiovisual object categorization. In primary auditory cortex, spatiotemporally aligned auditory and visual inputs are integrated into more salient units through automatic superadditive interactions. Based on our effective connectivity analysis, this auditory response amplification is most likely mediated via direct and indirect connectivity from visual cortices. In STS, IPS and PT, the profiles of audiovisual interactions were predicted by the subjects' behavioral benefits in categorization performance suggesting a role in integrating higher-order object features into perceptual representations. During explicit semantic categorization, the suppressive interactions in vlPFC may reflect multisensory facilitation of semantic retrieval, categorization and selection of an appropriate action. In conclusion, multisensory integration emerges at multiple processing stages and levels within the cortical hierarchy.

References

Avillac M, Ben Hamed S, Duhamel JR (2007) Multisensory integration in the ventral intraparietal area of the macaque monkey. J Neurosci 27:1922–1932.

Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J Cogn Neurosci 17:377–391.

Beauchamp MS (2005) Statistical criteria in fMRI studies of multisensory integration. Neuroinformatics 3:93–113.

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

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

Besle J, Fischer C, Bidet-Caulet A, Lecaignard F, Bertrand O, Giard MH (2008) Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. J Neurosci 28:14301–14310.

Brett M, Anton JL, Valabregue R, Poline JB (2002) Region of interest analysis using an SPM toolbox. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Neuroimage 16(2).

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.

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123.

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649–657.

Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage 14:427–438.

Calvert GA, Spence C, Stein BE (2004) The handbook of multisensory processes. Cambridge, MA: MIT.

Chao LL, Haxby JV, Martin A (1999) Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat Neurosci 2:913–919.

Dolan RJ, Fink GR, Rolls E, Booth M, Holmes A, Frackowiak RS, Friston KJ (1997) How the brain learns to see objects and faces in an impoverished context. Nature 389:596–599.

Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron 57:11–23.

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

Evans AC, Marrett S, Neelin P, Collins L, Worsley K, Dai W, Milot S, Meyer E, Bub D (1992) Anatomical mapping of functional activation in stereotactic coordinate space. Neuroimage 1:43–53.

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

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749–5759.

Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. Neuroreport 16:419–423.

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

Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC (1994)

Assessing the significance of focal activations using their spatial extent. Hum Brain Mapp 1:214–220.

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ (1995) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2:189–210.

Friston KJ, Holmes AP, Price CJ, Büchel C, Worsley KJ (1999) Multisubject fMRI studies and conjunction analyses. Neuroimage 10:385–396.

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19:1273–1302.

Friston KJ, Penny WD, Glaser DE (2005) Conjunction revisited. Neuroimage 25:661–667.

George N, Dolan RJ, Fink GR, Baylis GC, Russell C, Driver J (1999) Contrast polarity and face recognition in the human fusiform gyrus. Nat Neurosci 2:574–580.

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

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, Chandrasekaran C, Logothetis NK (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. J Neurosci 28:4457–4469.

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

Goebel R, van Atteveldt N (2009) Multisensory functional magnetic resonance imaging: a future perspective. Exp Brain Res 198:153–164.

Gosselin F, Schyns PG (2003) Superstitious perceptions reveal properties of internal representations. Psychol Sci 14:505–509.

Hein G, Doehrmann O, Müller NG, Kaiser J, Muckli L, Naumer MJ (2007) Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. J Neurosci 27:7881–7887.

Henson RN (2003) Neuroimaging studies of priming. Prog Neurobiol 70:53–81.

Holmes NP, Spence C, Hansen PC, Mackay CE, Calvert GA (2008) The multisensory attentional consequences of tool use: a functional magnetic resonance imaging study. PLoS One 3:e3502.

Kayser C, Logothetis NK (2009) Directed interactions between auditory and superior temporal cortices and their role in sensory integration. Front Integr Neurosci 3:7.

Kayser C, Petkov CI, Lippert M, Logothetis NK (2005) Mechanisms for allocating auditory attention: an auditory saliency map. Curr Biol 15:1943–1947.

Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional imaging reveals visual modulation of specific fields in auditory cortex. J Neurosci 27:1824–1835.

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

Kleiner M, Wallraven C, Bülthoff HH (2004) The MPI VideoLab—a system for high quality synchronous recording of video and audio from multiple viewpoints. MPI technical report 123.

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

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320:110–113.

Laurienti PJ, Burdette JH, Wallace MT, Yen YF, Field AS, Stein BE (2002) Deactivation of sensory-specific cortex by cross-modal stimuli. J Cogn Neurosci 14:420–429.

Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE (2005) On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. Exp Brain Res 166:289–297.

Lehmann C, Herdener M, Esposito F, Hubl D, di Salle F, Scheffler K, Bach DR, Federspiel A, Kretz R, Dierks T, Seifritz E (2006) Differential patterns of multisensory interactions in core and belt areas of human auditory cortex. Neuroimage 31:294–300.

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. J Neurosci 25:5148–5158.

Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. Trends Neurosci 28:264–271.

Macaluso E, Driver J, Frith CD (2003) Multimodal spatial representations engaged in human parietal cortex during both saccadic and manual spatial orienting. Curr Biol 13:990–999.

Martuzzi R, Murray MM, Michel CM, Thiran JP, Maeder PP, Clarke S, Meuli RA (2007) Multisensory interactions within human primary cortices revealed by BOLD dynamics. Cereb Cortex 17:1672–1679.

Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J Neurophysiol 56:640–662.

Miller LM, D'Esposito M (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci 25:5884–5893.

Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex 14:452–465.

Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K (2001) Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage 13:684–701.

Musacchia G, Schroeder CE (2009) Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. Hear Res 258:72–79.

Neal JW, Pearson RC, Powell TP (1990) The connections of area Pg, 7A, with cortex in the parietal, occipital and temporal lobes of the monkey. Brain Res 532:249–264.

Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005) Valid conjunction inference with the minimum statistic. Neuroimage 25:653–660.

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.

Noppeney U (2010) Characterization of multisensory integration with fMRI—experimental design, statistical analysis and interpretation. In: Frontiers in the neural bases of multisensory processes (Wallace MT, Murray MM, eds), in press. London: Taylor and Francis.

Noppeney U, Price CJ (2002) A PET study of stimulus- and task-induced semantic processing. Neuroimage 15:927–935.

Noppeney U, Price CJ, Penny WD, Friston KJ (2006) Two distinct neural mechanisms for category-selective responses. Cereb Cortex 16:437–445.

Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ (2008) The effect of prior visual information on recognition of speech and sounds. Cereb Cortex 18:598–609.

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Comparing dynamic causal models. Neuroimage 22:1157–1172.

Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004) Attentional modulation of human auditory cortex. Nat Neurosci 7:658–663.

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50:19–26.

Romanski LM (2004) Domain specificity in the primate prefrontal cortex. Cogn Affect Behav Neurosci 4:421–429.

Romanski LM (2007) Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. Cereb Cortex 17:i61–i69.

Sabsevitz DS, Medler DA, Seidenberg M, Binder JR (2005) Modulation of the semantic system by word imageability. Neuroimage 27:188–200.

Sadaghiani S, Maier JX, Noppeney U (2009) Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. J Neurosci 29:6490–6499.

Saito DN, Yoshimura K, Kochiyama T, Okada T, Honda M, Sadato N (2005) Cross-modal binding and activated attentional networks during audiovisual speech integration: a functional MRI study. Cereb Cortex 15:1750–1760.

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.

Schroeder CE, Smiley J, Fu KG, McGinnis T, O'Connell MN, Hackett TA (2003) Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. Int J Psychophysiol 50:5–17.

Seltzer B, Pandya DN (1994) Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus-monkey—a retrograde tracer study. J Comp Neurol 343:445–463.

Stanford TR, Quessy S, Stein BE (2005) Evaluating the operations underlying multisensory integration in the cat superior colliculus. J Neurosci 25:6499–6508.

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

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46:1004–1017.

Stevenson RA, James TW (2009) Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. Neuroimage 44:1210–1223.

Sugihara T, Diltz MD, Averbeck BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. J Neurosci 26:11138–11147.

Treisman M (1998) Combining information: probability summation and probability averaging in detection and discrimination. Psychol Methods 3:252–265.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289.

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

van Atteveldt N, Roebroeck A, Goebel R (2009) Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. Hear Res 258:152–164.

van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb Cortex 17:962–974.

Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RS (1996) Functional anatomy of a common semantic system for words and pictures. Nature 383:254–256.

Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. J Neurophysiol 76:1246–1266.

Wallace MT, Ramachandran R, Stein BE (2004) A revised view of sensory cortical parcellation. Proc Natl Acad Sci U S A 101:2167–2172.

Werner S, Noppeney U (2009) Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cereb Cortex. Advance online publication. Retrieved November 18, 2009. doi:10.1093/cercor/bhp248.

Wickens TD (2002) Multidimensional stimuli. In: Elementary signal detection theory, pp 172–194. New York: Oxford UP.

Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex 13:1034–1043.