Behavioral/Systems/Cognitive

Perceptual Decisions Formed by Accumulation of Audiovisual Evidence in Prefrontal Cortex

Uta Noppeney, Dirk Ostwald, and Sebastian Werner
Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

To form perceptual decisions in our multisensory environment, the brain needs to integrate sensory information derived from a common source and segregate information emanating from different sources. Combining fMRI and psychophysics in humans, we investigated how the brain accumulates sensory evidence about a visual source in the context of congruent or conflicting auditory information. In a visual selective attention paradigm, subjects (12 females, 7 males) categorized video clips while ignoring concurrent congruent or incongruent soundtracks. Visual and auditory information were reliable or unreliable. Our behavioral data accorded with accumulator models of perceptual decision making, where sensory information is integrated over time until a criterion amount of information is obtained. Behaviorally, subjects exhibited audiovisual incongruency effects that increased with the variance of the visual and the reliability of the interfering auditory input. At the neural level, only the left inferior frontal sulcus (IFS) showed an "audiovisual-accumulator" profile consistent with the observed reaction time pattern. By contrast, responses in the right fusiform were amplified by incongruent auditory input regardless of sensory reliability. Dynamic causal modeling showed that these incongruency effects were mediated via connections from auditory cortex. Further, while the fusiform interacted with IFS in an excitatory recurrent loop that was strengthened for unreliable task-relevant visual input, the IFS did not amplify and even inhibited superior temporal activations for unreliable auditory input. To form decisions that guide behavioral responses, the IFS may accumulate audiovisual evidence by dynamically weighting its connectivity to auditory and visual regions according to sensory reliability and decisional relevance.

Introduction

Selecting an appropriate action on the basis of unreliable sensory information is one of the most fundamental cognitive tasks facing the brain. To form a perceptual decision, the brain is thought to accumulate noisy sensory information over time until a decisional threshold is reached (Mazurek et al., 2003; Schall, 2003; Smith and Ratcliff, 2004; Lo and Wang, 2006; Gold and Shadlen, 2007). Neurophysiological studies have demonstrated neural activity reflecting an evolving decision in areas associated with motor planning and attention, including lateral intraparietal area (LIP) (Kiani et al., 2008), ventral premotor (Romo et al., 2004) and dorsolateral prefrontal (Kim and Shadlen, 1999) cortices. Most prominently, during visual motion discrimination, the neuronal firing rate in LIP builds up progressively until a critical firing rate is reached, the decision process terminated, and a response elicited. The accumulation rate of the neuronal activity is proportional to the amount of the sensory evidence. The response times (RTs) as indices for the time to decisional threshold decrease with increasing sensory evidence.

Given the limited temporal resolution of the blood oxygenation level-dependent (BOLD) signal, human fMRI studies have identified candidate "accumulator" regions primarily based on their correlations between trial-to-trial variations in BOLD response and subjects' response times. Unisensory, e.g., auditory (Binder et al., 2004), visual (Heekeren et al., 2004; Grinband et al., 2006; Thielscher and Pessoa, 2007), or somatosensory (Pleger et al., 2006), decisions were associated with activations in anterior cingulate and dorsolateral prefrontal cortices. Yet, under natural conditions, multiple senses are simultaneously excited by inputs that emanate from common or different sources in the environment (Schroeder and Foxe, 2005; Ghazanfar and Schroeder, 2006; Driver and Noesselt, 2008). For instance, running through the forest, catching sight of a wild boar, we may concurrently hear its grunts or unrelated birdsong. Ideally, the human brain should integrate sensory information derived from a common source, while avoiding mergence of information from different sources. Indeed, multisensory integration breaks down, when sensory estimates are brought into conflict. Nevertheless, even when no coherent multisensory percept can be formed, conflicting information from one sensory modality interferes with decisions on sensory inputs from another modality as shown in selective attention paradigms. The present study used a visual selective attention paradigm to focus on audiovisual interactions at the decisional level and investigate how (in)congruent auditory information interacts with accumulation of visual evidence during object categorization. Subjects categorized visual action movies while ignoring the semantically congruent or incongruent soundtracks that were always presented with spatiotemporally coincident onsets. Auditory and visual signals were reliable or unreliable. Given subjects' lifelong exposure to environmental

Received Jan. 20, 2010; revised April 7, 2010; accepted April 16, 2010.

This work was supported by the Deutsche Forschungsgemeinschaft and the Max-Planck Society. We thank Peter Dayan and Richard Lewis for helpful discussions, Mario Kleiner for advice with recording the stimuli, and Beatrix Barth for help with acquiring the psychophysics data.

Correspondence should be addressed to Uta Noppeney, Max Planck Institute for Biological Cybernetics, Spemannstrasse 41, 72076 Tübingen, Germany. E-mail: uta.noppeney@tuebingen.mpg.de.

DOI:10.1523/JNEUROSCI.0455-10.2010

statistics, we reasoned that subjects a priori expected the auditory and visual signals to emanate from a common source, leading to stronger audiovisual interference effects at trial onset and protracted evidence accumulation for incongruent trials. Based on the compatibility bias model (Yu et al., 2009), these incongruency effects should decrease with the reliability of the visual input (to be categorized), yet increase with the reliability of the interfering auditory input. Thus, the accumulation process should be slowest for trials with unreliable visual and incongruent reliable auditory information, resulting in (1) greater BOLD responses in multisensory "accumulator" regions and (2) longer response times. From the perspective of functional integration, we hypothesized that evidence accumulation and resolution of audiovisual (in)congruency relies upon recurrent message passing among hierarchically arranged cortical areas as described in predictive coding schemes. In predictive coding formulations of hierarchical inference, backward connections furnish the top-down predictions and forward connections the prediction errors that correspond to the bottom-up (sensory) evidence that has yet to be explained by the top-down predictions. Using dynamic causal modeling, we therefore investigated whether sensory reliability and audiovisual incongruency modulated the forward or backward connections between potential accumulator and sensory areas.

Materials and Methods

Subjects

After giving informed consent, 19 healthy volunteers (12 females, 18 right-handed; mean age 22.1 years, range 19–26 years) participated in the fMRI study and 10 different healthy volunteers (6 females, 9 right-handed; mean age 25.5 years, range 21–39 years) in the additional psychophysics study outside the scanner. All subjects had normal or corrected-to-normal vision and reported normal hearing. The study was approved by the human research review committee of the University of Tübingen.

Stimuli

Stimuli were grayscale 2 s video clips and the corresponding sounds of actions associated with 15 tools (e.g., hammer) and 15 musical instruments (e.g., violin) recorded at the MPI-VideoLab (Kleiner et al., 2004). The actor's hands were included in the video clips. The two distinct categories were selected to allow for a semantic categorization task. However, category-selective activations are not the focus of this communication (Chao et al., 1999; Lewis et al., 2004, 2005; Noppeney et al., 2006; Stevenson and James, 2009).

Reliability of the images was manipulated by applying different degrees of Fourier phase scrambling. To this end, original movie frames (i.e., tools and musical instruments) and uniform random noise images were separated into spatial frequency amplitude spectra and phase components using the Fourier transform. Two levels of visual reliability were generated by combining the original amplitude spectra with (1) the original phase components (i.e., intact vision) or (2) phase components representing a linear interpolation between original and random noise phase spectra (i.e., degraded vision). The linear interpolation preserved 20% of the original phase components. Based on initial piloting, this level was selected to maximize the incongruency effect in terms of reaction times when accuracy was emphasized in the task instructions (see below). The phase randomization procedure ensured that movie frames at both levels of reliability were matched in terms of their spatial frequency content, distribution of phase components, and low-level statistics [i.e., mean luminance and root-mean-square (RMS) contrast] (Dakin et al., 2002). To prevent subjects from using low-level visual cues for categorization, we selected and matched the mean movie frames with respect to their mean luminance ($t_{(28)} = 0.4971$; $p = 0.3115$) and RMS contrast ($t_{(28)} = 1.2298$; $p > 0.1145$).

Auditory stimuli were the sounds produced by the actions of the tools and musical instruments during the recording of the video clips. Each sound file (2 s duration, 48,000 Hz sampling rate) was equated for maximum/minimum intensity of the sound stimulation. Similar to the visual domain, original and white noise sounds were transformed into Fourier amplitude and phase components. Two levels of auditory reliability were generated by combining the original temporal frequency amplitude spectra with (1) the original phase components (i.e., intact sound) or (2) phase components representing a linear interpolation between original and white noise phase spectra (i.e., degraded audition). The linear interpolation between original and white noise phase spectra preserved 30% of the original phase components. The procedure ensured that sounds across the two levels of reliability were matched in terms of their temporal frequency contents, distribution of phase components, and RMS power. The sounds from the two categories (i.e., tools or musical instruments) were matched with respect to their RMS power ($p > 0.05$). Each 2 s sound file was presented monophonically and concurrently with the presentation of the 2 s video clip.

Video frames and sounds were recombined into semantically congruent and incongruent movie clips using Adobe Premiere Pro 2.0 software (Adobe Systems). Incongruent stimuli combined a video of a musical instrument and a sound of a tool and vice versa. To control for stimulus effects, each auditory or visual stimulus component was combined into eight incongruent audiovisual (AV) movies that were rotated over the four incongruent conditions. In this way, congruent and incongruent conditions were equated with respect to the auditory and visual inputs and only differed in the relationship (i.e., semantically congruent vs incongruent) between the auditory and visual components.

Experimental design

In a visual selective attention paradigm, subjects were presented with audiovisual movies of hand actions that involved tools or musical instruments [for auditory selective attention paradigm, see Noppeney et al. (2008)]. In a two-alternative forced-choice task, they categorized the video clips as tools or musical instruments while ignoring the concurrent congruent or incongruent sound tracks. The video clips and the concurrent source sounds were either (1) semantically congruent (auditory and visual inputs emanated from the same object, e.g., a video of a violin paired in synchrony with the sound produced by the violin) or (2) semantically incongruent (auditory and visual inputs emanated from objects of opposite categories, e.g., a video of a violin paired with a hammering sound). Both auditory and visual information could be intact (reliable) or degraded (unreliable). Hence, the 2 × 2 × 2 factorial design manipulated the following: (1) visual reliability (intact = V, degraded = v), (2) auditory reliability (intact = A, degraded = a), and (3) semantic incongruency of the video clips and the soundtracks (congruent = C, incongruent = I) (Fig. 1). In all conditions (i.e., semantically congruent and incongruent), auditory and visual inputs emanated from matched spatial locations (i.e., sound and videos were presented centrally). Furthermore, in all conditions, auditory and visual inputs were presented synchronously with respect to stimulus onsets. Yet, since the time courses of actions/sounds from different objects were not temporally matched, the semantically incongruent stimuli induced audiovisual asynchrony over the 2 s duration of the movies. In contrast, congruent stimuli were always synchronous. Hence, (in)congruency included two components, semantic incongruency and temporal asynchrony (with respect to the time course, but not the onset, of the auditory and visual signals).

On each trial the stimulus was presented for 2 s followed by 800 ms of fixation (i.e., stimulus onset asynchrony of 2800 ms). Subjects responded as quickly and accurately as possible during the 2 s stimulus presentation period. The mapping from stimulus category to button/finger was counterbalanced across subjects. Fifty percent of the trials required a "tool" response. Fifty percent of the trials were semantically congruent. The stimulus duration did not depend on subjects' response time, i.e., the stimulus was not terminated based on subjects' response but fixed to 2 s. A fixed duration of stimulus presentation was used, so that the variation in BOLD response attributable to response processing was not confounded by variation in stimulus duration. Thus, the experimental paradigm combined (1) fixed stimulus duration and (2) speed–accuracy instructions. The speed–accuracy trade-off was manipulated across the psychophysics and the fMRI study using two different task instructions.

(1) In the psychophysics study, the task instructions emphasized response speed rather than accuracy to obtain categorization accuracy at different response time bins and provide insights into the within-trial dynamics of subjects' beliefs about the category of the visual object. (2) In the fMRI study, we emphasized accuracy rather than speed to encourage subjects to gather information about visual object's category to a high level of certainty. Unless otherwise stated, identical parameters were used in the psychophysics and fMRI study.

Psychophysics study. In the psychophysics study (outside the scanner), emphasis was placed on response speed rather than accuracy. Subjects were instructed to accept a reduction in response accuracy to maximize response speed. To obtain sufficient fast and error responses, they were given feedback on their accuracy level every 15 trials. They were encouraged to (1) respond faster for accuracy >0.75, (2) respond slower and more accurately for accuracy <0.6, and (3) keep going for 0.6 ≤ accuracy ≤ 0.75. After initial training, subjects participated in eight sessions. For comparison with the fMRI study and to limit learning effects, only the first two sessions were included in the current study. The six additional sessions were acquired to characterize learning effects and will be the focus of a future communication. In each session, each of the 30 stimuli (15 musical instruments and 15 tools) was presented once in each of the eight conditions, amounting to 240 trials per session (i.e., 30 × 8). Each subject was presented a particular incongruent audiovisual stimulus combination only once within the first two sessions to prevent subjects from learning new incongruent associations.

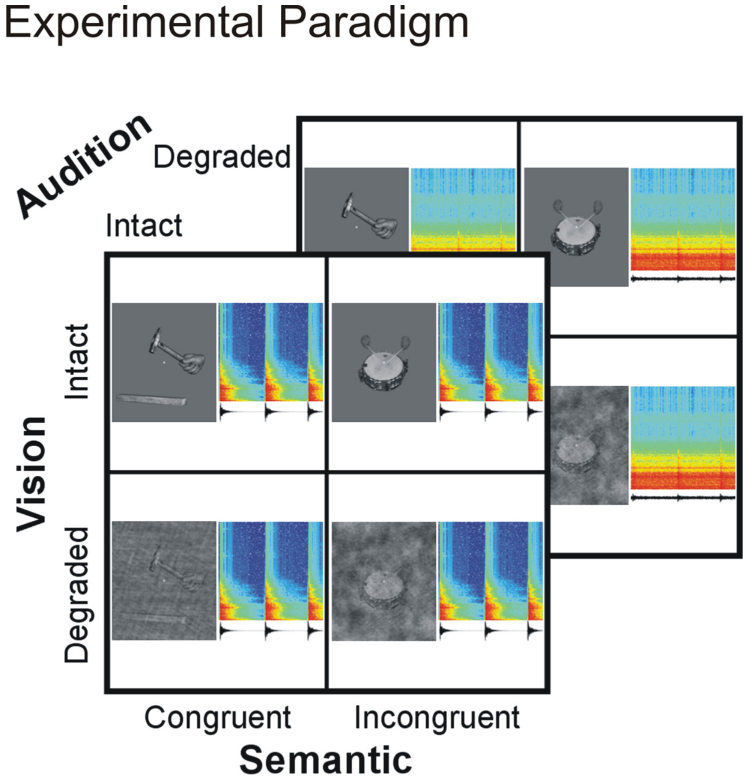

Figure 1. Experimental paradigm. In a visual selective attention paradigm, subjects categorized the video clips as tools (e.g., hammer) or musical instruments (e.g., drum) while ignoring the concurrent congruent or incongruent sounds. The 2 × 2 × 2 factorial design manipulated the following: (1) visual reliability: intact versus degraded video clips; (2) auditory reliability: intact versus degraded sounds; and (3) semantic (in)congruency: incongruent versus congruent. For each condition, an audiovisual stimulus pair is represented by an image of the video clip and the waveform and time frequency spectrogram of the sound.

fMRI study. In the fMRI study (inside the scanner), subjects responded during the 2 s stimulus presentation period as accurately and quickly as possible. Importantly, a special emphasis was placed on accuracy rather than speed, so that the conditions differed primarily in response time rather than accuracy (though we acknowledge that this was not perfectly achieved). After initial training, subjects participated in two sessions inside the scanner. In each session, each of the 30 stimuli (15 musical instruments and 15 tools) was presented once in each of the eight conditions, amounting to 240 trials per session (i.e., 30 × 8). Each subject was presented a particular incongruent audiovisual stimulus combination only once to prevent subjects from learning new incongruent associations. Blocks of eight activation trials (block duration ~23 s) were interleaved with 8 s fixation. To maximize design efficiency, a pseudorandomized sequence of stimuli and activation conditions was generated for each subject.

Experimental rationale, compatibility bias model, and expected response profile

The current study introduced audiovisual incongruency to attenuate AV integration processes, leading to a coherent multisensory percept and focus selectively on AV interactions at the "decisional" level. To provide a normative Bayesian perspective on the interfering effect of task-irrelevant auditory input on visual perceptual decisions and its within-trial dynamics, we adapted the "compatibility bias model" (Yu et al., 2009). Applied to the multisensory context of our visual selective attention paradigm, the basic idea of the "compatibility bias model" is that—as a result of lifelong adaptation to the statistics of the natural environment—humans have developed prior expectations of auditory and visual signals being congruent (i.e., emanate from a common source) when they co-occur in space and time as in the current experiment. During the course of each trial, subjects accumulate evidence concomitantly about (1) the "true" (i.e., congruent or incongruent) relationship of the auditory and visual signals and (2) the category (tool vs musical instrument) of the visual object. The interference of incongruent auditory information on the accumulation of visual object evidence should then be particularly pronounced at trial onset when subjects' congruency prior dominates and decreases during the course of the trial, when incoming evidence overrides these prior expectations. The accumulation process is terminated when the evidence about the visual object category reaches a decisional threshold and the subject "opts for" one of the two alternatives (i.e., tool vs musical instrument) [for further details on the implementation of the model, see supplemental material (available at www.jneurosci.org) and Yu et al. (2009); for relationship to drift diffusion model, see Liu et al. (2009)].

Based on the temporal dynamics of evidence accumulation, we expect the following characteristic profile for (1) accuracy versus response time functions (Servan-Schreiber et al., 1998a,b), (2) response times, and (3) BOLD response profile across the eight conditions in our 2 × 2 × 2 factorial design. (1) Under "speed" instructions (i.e., psychophysics experiment), we expect accuracy versus response time functions to diverge for incongruent and congruent trials progressively with decreasing response times (Fig. 2*A*). This is because the congruency prior induces interference primarily at trial onset, leading to a dip in accuracy even below chance for incongruent trials with fast response times. (2) Particularly under accuracy instructions (i.e., fMRI experiment), the temporal dynamics of evidence accumulation leads to a characteristic profile of the condition-specific times to decisional threshold as indexed by subjects' reaction times (Fig. 2*B*). Since the degradation of the visual information delays inference about both visual object category and audiovisual (in)congruency, auditory interference effects are more pronounced for visual unreliable than reliable conditions. Conversely, auditory degradation reduces the interference effect of incongruent auditory input. (3) Given the proposed links between evidence accumulation and the rise in

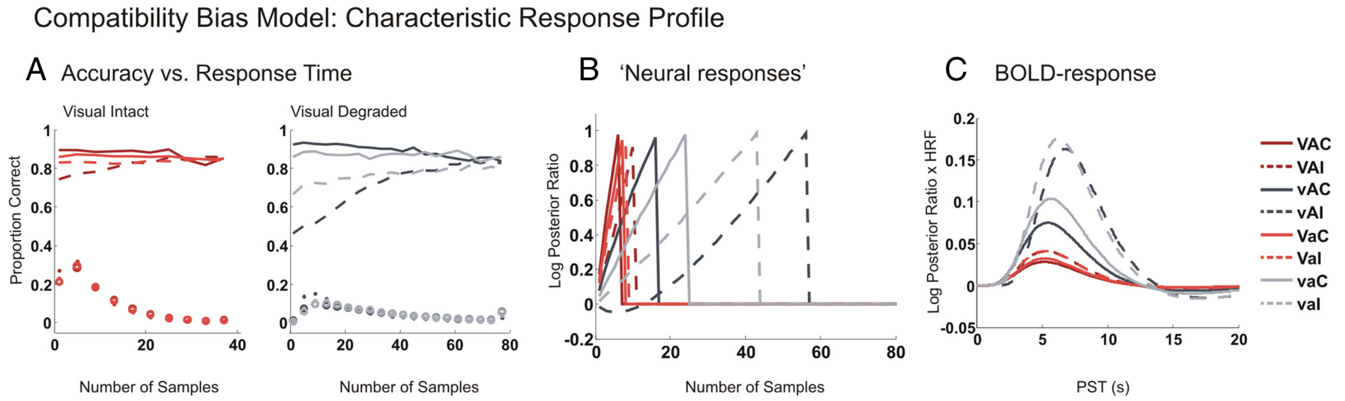

Figure 2. Compatibility bias model and the expected characteristic response profile. Predictions of the compatibility bias model averaged across 30,000 simulated trials with 100 sampling points (for details, see supplemental material, available at www.jneurosci.org). Mean trajectories of accuracy (***A***), log posterior ratio (***B***), and predicted BOLD responses (***C***) are displayed for the eight conditions in the 2 × 2 × 2 factorial design. The conditions are coded in color. V, Intact vision; v, degraded vision; A, intact audition; a, degraded audition; C, congruent; I, incongruent. ***A***, Accuracy as a function of model RT (dashed and solid lines) and the RT distribution (circles and dots). The difference in accuracy between congruent (solid) and incongruent (dashed) trials is amplified for trials with short response times. ***B***, The log posterior ratio (favoring the correct over the wrong visual category) has been related to neural responses in brain regions involved in evidence accumulation. In these simulations, it was assumed that subjects accumulate evidence for one visual category to a fixed threshold. The time to threshold indexes observer's decision time and is related to his response time. ***C***, The simulated "neural responses" (***B***) were convolved with the hemodynamic response function to obtain ordinal predictions of the BOLD response separately for each of the eight conditions. The predicted BOLD responses differ primarily in magnitude. Most notably, the incongruency effects for both response times and BOLD responses are increased when the visual stimulus is unreliable and the auditory stimulus is reliable.

neural activity in putative "accumulator regions," the log posterior ratio favoring one response over another may be used as an index for neural activity (Fig. 2*B*). Since subjects were instructed to respond as accurately and fast as possible, we assumed that the neural activity rises to a decisional threshold, when a response is elicited and the neural activity returns back to baseline as described in the reaction time paradigms of neurophysiological studies (Mazurek et al., 2003). Even though fMRI can obviously not characterize the fine-grained within-trial temporal dynamics of the accumulation process (Philiastides and Sajda, 2006; Kaiser et al., 2007; de Lange et al., 2010), under the assumption of a linear convolution model, we can convolve the predicted ramps of the neuronal activity for the eight conditions with the hemodynamic response function to generate an expected BOLD response profile of a putative multisensory decision region. However, we did not use this model to quantitatively predict condition-specific BOLD responses, but simply to establish the nature and direction of the interactions between auditory/visual reliability and AV (in)congruency. As illustrated in Figure 2, *B* and *C*, the BOLD response in a "multisensory accumulator region" and reaction times are expected to show (1) incongruency effects that increase with the variance of the visual input [interaction between (in)congruency and visual reliability] and (2) incongruency effects that decrease with the variance of the auditory input [interaction between (in)congruency and auditory reliability]. In line with classical interaction analyses for identification of "low level" spatiotemporal or perceptual audiovisual integration processes (Calvert et al., 2001; Noppeney, 2010), our factorial design enables us to reveal audiovisual integration processes at the decisional level by the interaction between congruency and visual (or auditory) reliability.

Experimental setup

Visual and auditory stimuli were presented using the Cogent Toolbox (John Romaya, Vision Lab, UCL; www.vislab.ucl.ac.uk) running under MATLAB (MathWorks).

Auditory stimuli were presented at ~80 dB SPL, using MR-compatible headphones (MR Confon). Visual stimuli (size 5.7° × 5.7° visual angle) were back-projected onto a Plexiglas screen using a LCD projector (JVC) visible to the subject through a mirror mounted on the MR head coil. Subjects performed the behavioral task using a MR-compatible custom-built button device connected to the stimulus computer.

MRI data acquisition

A 3T Siemens Magnetom Trio System was used to acquire both T1-weighted anatomical images and T2*-weighted axial echoplanar images with BOLD contrast [gradient echo, 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, voxel size = 3.0 mm × 3.0 mm × (2.6 mm slice thickness + 0.4 mm interslice gap)]. There were two sessions with a total of 320 volume images per session. The first six volumes (except for five volumes in one subject) were discarded to allow for T1 equilibration effects. A three-dimensional high-resolution anatomical image was acquired (TR = 10.55 ms, TE = 3.14 ms, TI = 680 ms, flip angle = 22°, FOV = 256 mm × 224 mm × 176 mm, image matrix = 256 × 224 ×176, isotropic spatial resolution 1 mm).

Data analysis

fMRI data analysis. The functional MRI data were analyzed with statistical parametric mapping [SPM from the Wellcome Department of Imaging Neuroscience, London; 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 (Talairach and Tournoux, 1988; Evans et al., 1992), resampled to 3 × 3 × 3 mm^3 voxels, and spatially smoothed with a Gaussian kernel of 8 mm FWHM. The time series in each voxel was high-pass filtered to 1/128 Hz. The fMRI experiment was modeled in an event related fashion 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 eight conditions in our 2 × 2 × 2 factorial design, the statistical model included missed responses as a separate condition. As error trials were shown to produce similar neuronal activity as correct trials, we pooled correct trials and errors (Mazurek et al., 2003). Nuisance covariates included the realignment parameters (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. This involved creating eight contrast images (i.e., each of the eight conditions summed over the two sessions) for each subject and entering them into a second-level ANOVA.

Inferences were made at the second level to allow a random-effects analysis and inferences at the population level (Friston et al., 1995). Unless otherwise stated, we report activations at $p < 0.05$ at the cluster level corrected for multiple comparisons within the neural systems activated relative to fixation (at $p < 0.001$ uncorrected) using an auxiliary (uncorrected) voxel threshold of $p < 0.001$. This auxiliary threshold defines the spatial extent of activated clusters, which forms the basis of our (corrected) inference.

Effective connectivity analysis: dynamic causal modeling. Dynamic causal modeling (DCM) treats the brain as a dynamic input–state–output system (Friston et al., 2003). 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, which characterize effective connectivity among regions; and (3) modulatory parameters, which characterize changes in effective connectivity caused by experimental manipulation. This third set of parameters, the modulatory effects, allows us to explain fMRI incongruency or sensory reliability effects 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 under complexity constraints. This adjustment corresponds to the model fitting.

For each subject, 24 DCMs (Friston et al., 2003) were constructed. Each DCM included three regions: (1) the left inferior frontal sulcus as an "accumulator" region (IFS; $x = -54, y = 15, z = 33$), (2) a right fusiform region that showed increased activation for incongruent relative congruent stimuli (FFG; $x = 36, y = -45, z = -15$), (3) a left superior temporal region that was activated for all stimuli > baseline (STG; $x = -45$, $y = -15$, $z = 0$) (see Fig. 6*A*, left). The right FFG was chosen as the visual input region as this was functionally associated with the representation of object information. Hence, it may be a candidate region for providing object evidence for the left IFS. Similarly, the left STG showed increased activation for intact relative to degraded auditory object stimuli and may thus be involved in representing auditory object evidence. The three regions were bidirectionally connected. The timings of the onsets were individually adjusted for each region to match the specific time of slice acquisition. Visual stimuli were entered as extrinsic inputs to FFG and auditory stimuli to STG. Holding the number of parameters and the intrinsic and extrinsic connectivity structure constant, the $24 = 2 \times 2 \times 6$ DCMs factorially manipulated the connection that was modulated by the three main effects: (1) visual reliability modulated the forward versus backward connection between FFG and IFS, (2) auditory reliability modulated the forward versus backward connection between STG and IFS, and (3) the effect of AV incongruency affected any one of the six connections. Each effect was allowed to modulate exactly one connection in a particular DCM (see Fig. 6*A*).

The regions were selected using the maxima of the relevant contrasts from our random-effects analysis. Region-specific time series (concatenated over the two sessions and adjusted for confounds) comprised the first eigenvariate of all voxels within a 4-mm-radius sphere centered on the subject-specific peak in the relevant contrast. The subject-specific peak was uniquely identified as the maximum within the relevant contrast in a particular subject in a 9-mm-radius sphere centered on the peak coordinates from the group random-effects analysis.

Bayesian model comparison. To determine the most likely of the 24 DCMs given the observed data from all subjects, we implemented a fixed- (Penny et al., 2004) and a random- (Stephan et al., 2009) effects group analysis. The fixed-effects group analysis was implemented by taking the product of the subject-specific Bayes factors over subjects (this is equivalent to the exponentiated sum of the log model evidences of each subject-specific DCM) (Penny et al., 2004). In brief, given the measured data y and two competing models, Bayes factors are the ratio of the evidences of the two models (Kass and Raftery, 1995). A Bayes factor of one represents equal evidence for the two models. A Bayes factor above 3 is considered positive evidence for one of the two models. The model evidence as approximated by the free energy depends not only on model fit but also model complexity. Here, we have limited ourselves to the 24 models that were equated for the number of parameters. 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 (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 (1) the expectation of this posterior probability, i.e., the expected likelihood of obtaining the *k*th model for any randomly selected subject, and (2) 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, which 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 modulatory, extrinsic, and intrinsic connection strengths were entered into *t* tests at the group level. This allowed us to summarize the consistent findings from the subject-specific DCMs using classical statistics.

Model comparison and statistical analysis of connectivity parameters of the optimal model enables us to address the following two questions. First, from the perspective of hierarchical Bayesian inference where evidence accumulation relies on recurrent message passing between multiple cortical hierarchical levels, we investigated how visual and auditory reliabilities influence evidence accumulation via distinct modulations of forward or backward connections between the IFS and sensory areas. As the visual selective attention paradigm renders visual information task-relevant and auditory information task-irrelevant or even interfering, we hypothesized that the connections from IFS to visual and auditory areas may be modulated asymmetrically. Second, we examined whether the incongruency effects in the right fusiform are mediated via connections from STG or backwards connections from IFS.

Results

In the following, we report (1) the behavioral results from the psychophysics study (outside the scanner), (2) the behavioral results from the fMRI study, (3) the fMRI results of the conventional analysis focusing on regionally selective activations, and (4) the DCM results providing insights into potential neural mechanisms that mediate the observed regional activations.

Behavioral results—psychophysics study

For performance accuracy, a three-way ANOVA with visual reliability (intact vs degraded), auditory reliability (intact vs degraded), and congruency (congruent vs incongruent) identified significant main effects of congruency ($F_{(1,9)} = 14.5$; $p = 0.004$) and visual reliability ($F_{(1,9)} = 64.1$; $p < 0.001$). In addition, there was a significant interaction effect between congruency and visual reliability ($F_{(1,9)} = 12$; $p = 0.007$).

For reaction times, a three-way ANOVA identified significant main effects of congruency ($F_{(1,9)} = 36$; $p < 0.001$) and visual reliability ($F_{(1,9)} = 104.3$; $p < 0.001$). Reaction times were longer for visual degraded than intact trials and for incongruent than congruent trials. Furthermore, interactions were observed between (1) congruency and visual reliability ($F_{(1,9)} = 4.2$; $p = 0.07$) and (2) congruency and auditory reliability ($F_{(1,9)} = 14.8$; $p = 0.004$). More specifically, degraded vision increased the incongruency effect, while degraded audition decreased the incongruency effect (see supplemental Table 1, available at www.jneurosci.org as supplemental material).

To characterize the within-trial dynamics of subjects' beliefs about the category of the visual object, subjects' categorization responses were sorted according to reaction times and assigned to equally spaced reaction time bins of 100 ms (except for the first

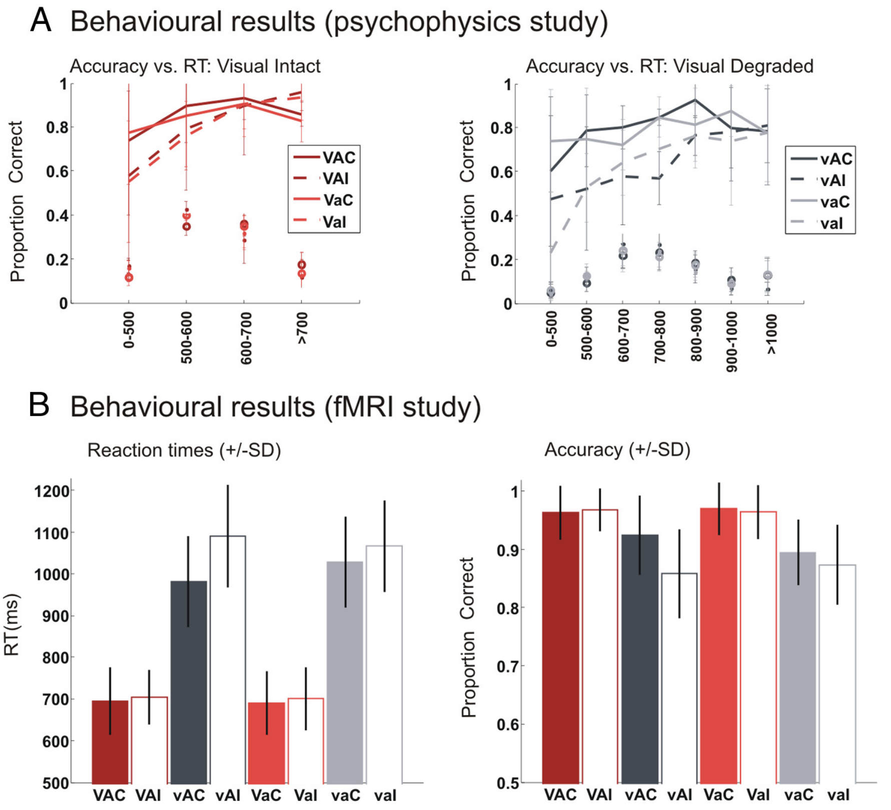

Figure 3. Behavioral results. ***A***, Psychophysics study outside the scanner: accuracy (across-subjects mean ± SD) as a function of reaction time bin is displayed for visual intact (left) and visual degraded (right) conditions. The dashed and solid lines denote the empirical probability of making a correct response as a function of binned RT; the markers (dots and circles) indicate the empirical distribution of response times over RT bins. The difference in the probability of correct responses between congruent (solid) and incongruent (dashed) conditions diverges for short response times. This profile is observed consistently for visual intact and degraded conditions, though the exact shape of the curves differ. Thus, as predicted by the compatibility bias model, the incongruency effect is most pronounced for trials with short response times. ***B***, Behavioral performance during the fMRI study: response times (left) and accuracy (right) for the eight conditions (across-subjects means ± SD). V, Intact vision; v, degraded vision; A, intact audition; a, degraded audition; C, congruent; I, incongruent.

and last reaction time bins, which included all trials with response times below or above a given threshold to allow for reliable estimation). To ensure that the bins for short response times were similarly influenced by all subjects, subjects' response times were normalized to a common group mean (by subtracting the difference between the response time median for a particular subject and the group mean of subjects' medians from all response times within that subject). Since the empirical distribution over reaction time bins was more spread out for degraded relative to intact conditions (Fig. 3*A*, dots), the response times were assigned to four bins ($<$500 ms, 500 $\leq x <$ 600 ms, 600 $\leq x <$ 700 ms, $\geq$700 ms) for intact conditions and to six bins ($<$500 ms, 500 $\leq x <$ 600 ms, 600 $\leq x <$ 700 ms, 700 $\leq x <$ 800 ms, 800 $\leq x <$ 900 ms, 900 $\leq x <$ 1000 ms, $\geq$1000 ms) for degraded conditions. Figure 3*A* shows the empirical distributions over reaction time bins (dots, circles) and the empirical probability of making a correct response as a function of binned RT (dashed and solid lines). In both intact (left) and degraded (right) conditions, the congruent (solid lines) and incongruent (dashed lines) curves diverge for short response times. In contrast, the accuracy levels for long response times are comparable across congruent and incongruent conditions. This impression was validated statistically in three-way ANOVAs of response accuracy, performed separately for visual intact and degraded conditions (all results are reported Greenhouse–Geisser corrected, missing values were replaced by across-subjects mean). For visual intact conditions, a three-way ANOVA with auditory reliability (intact vs degraded), congruency (congruent vs incongruent), and reaction time bin (four levels) identified a significant main effect of reaction time bin ($F_{(2.1,18.6)} = 18.4$; $p < 0.001$) and a significant interaction between congruency and reaction time bin ($F_{(2.5,22.9)} = 8.7$; $p = 0.001$). For visual degraded conditions, a three-way ANOVA with auditory reliability (intact vs degraded), congruency (congruent vs incongruent), and reaction time bin (seven levels) identified significant main effects of congruency ($F_{(1,9)} = 16.1$; $p = 0.003$) and reaction time bin ($F_{(2.2,19.7)} = 10.2$; $p = 0.001$) and a significant interaction between congruency and reaction time bin ($F_{(3.6,32.7)} = 4.5$; $p = 0.006$). The significant interaction between congruency and reaction time bin indicates that the effect of incongruent auditory information is particularly pronounced for short response times as predicted by the compatibility bias model.

Behavioral results—fMRI study

For performance accuracy, a three-way ANOVA with visual reliability (intact vs degraded), auditory reliability (intact vs degraded), and congruency (congruent vs incongruent) identified significant main effects of congruency ($F_{(1,18)} = 30.3$; $p < 0.001$) and visual reliability ($F_{(1,8)} = 79.3$; $p < 0.001$). In addition, there was a significant interaction effect between congruency and visual reliability ($F_{(1,18)} = 15.1$; $p = 0.001$) and a significant three-way interaction ($F_{(1,18)} = 9.4$; $p = 0.007$). Thus, even though task instructions placed more emphasis on accuracy than speed, subjects' speed–accuracy trade-off still led to differences in accuracy across conditions.

For reaction times, a three-way ANOVA identified significant main effects of congruency ($F_{(1,18)} = 42.2$; $p < 0.001$) and visual reliability ($F_{(1,8)} = 440.1$; $p < 0.001$). Reaction times were longer for visual degraded than intact trials and for incongruent than congruent trials. Crucially, there were significant interactions between (1) congruency and visual reliability ($F_{(1,18)} = 27.6$; $p < 0.001$) and (2) congruency and auditory reliability ($F_{(1,18)} = 17.9$; $p = 0.001$). More specifically, degraded vision increased the incongruency effect, while degraded audition decreased the incongruency effect. Thus, as predicted by the compatibility bias model, visual and auditory reliability exerted opposite effects on the incongruency effects. Furthermore, we observed a significant three-way interaction ($F_{(1,18)} = 10.1$; $p = 0.005$) (see supplemental Table 2, available at www.jneurosci.org as supplemental material, and Fig. 3*B*).

A further characterization of subjects' accuracy as a function of response times was not applied to the behavioral data from the fMRI study, since the task instructions primarily emphasizing accuracy rather than response speed did not provide us with a sufficiently widespread response distribution.

Table 1. Effects of visual and auditory reliability

Region	Coordinates			*z*-score peak	Number of voxels	*p* value (cluster) corrected*
Visual: intact > degraded						
R. lateral occipital sulcus	36	−87	−9	6.1	62	0.005
L. lateral occipital sulcus	−36	−87	−9	5.7	72	0.002
Visual: degraded > intact						
L. occipital pole/cuneolingual gyrus	−9	−96	−3	>8	2814	<0.001
R. occipital pole	15	−99	3	>8		
R. lat. occipitotemporal sulcus	45	−48	−18	5.8		
L. lat. occipitotemporal sulcus	−45	−51	−12	5.9	102	<0.001
R. intraparietal sulcus	27	−60	45	6.8		
L. intraparietal sulcus	−30	−51	42	7.8		
L. posterior middle temporal gyrus	−57	−51	9	4.8	75	0.002
R. inferior frontal/precentral sulcus	42	6	33	>8	523	<0.001
L. inferior frontal/precentral sulcus	−39	9	27	>8	463	<0.001
Auditory: intact > degraded						
R. superior temporal gyrus/Heschl's gyrus	63	−21	6	7.7	1156	<0.001
L. superior temporal gyrus/Heschl's gyrus	−51	−33	12	7.5	919	<0.001
R. inferior frontal gyrus	57	24	30	4.3	37	0.028

*Volume of interest = 11,334 voxels activated for stimulus > fixation at $p < 0.001$, extent threshold >20 voxels. L., Left; lat., lateral; R., right.

Table 2. Incongruency effects and their interactions with visual or auditory reliability

Region	Coordinates			*z*-score peak	Number of voxels	*p* value (cluster) corrected*
Interaction: I > C for v > V						
L. Inf. frontal/precentral sulcus	−54	12	33	4.2	82	0.001
R. intraparietal sulcus	24	−57	51	3.5	4	NS
Interaction: I > C for A > a						
L. Inf. frontal/precentral sulcus	−51	21	33	3.4	8	NS
R. intraparietal sulcus	27	−69	39	3.2	5	NS
Incongruent > congruent						
R. fusiform gyrus	36	−45	−15	4.8	56	0.007
R. intraparietal sulcus	30	−72	30	3.8	34	0.036
R. posterior superior temporal sulcus	45	−81	18	3.7	35	0.033

*Volume of interest = 11,334 voxels activated for stimulus > fixation at $p < 0.001$, extent threshold >20 voxels. Inf., Inferior; L., left; R., right.

Summary of the behavioral results

In line with the compatibility bias model, the accuracy response time functions diverged for incongruent and congruent trials with decreasing response times. This temporal profile resulted from a dip in accuracy for incongruent trials with short response times. Furthermore, in terms of response times, the incongruency effects increased with the variance of the visual input and the reliability of the auditory input.

fMRI results

Main effect of visual or auditory reliability

Candidate regions that are involved in representing auditory or visual object evidence were identified by directly comparing activations induced by intact and degraded stimuli in the visual or auditory modalities (Table 1). While intact relative to degraded auditory stimuli increased activations in the bilateral superior temporal gyri spreading from posterior portions (planum temporale) to the anterior STG, no areas showed increased activations for degraded relative to intact auditory stimuli. Thus, reliability of the irrelevant auditory signal increased activations in the auditory processing system. The response profile in the auditory domain sharply contrasts with that observed for the effect of visual reliability. Here, degraded relative to intact visual stimuli increased activations in a widespread bilateral neural system encompassing the occipital pole, lateral occipitotemporal, intraparietal, and inferior frontal/precentral sulci. Intact relative to degraded visual stimuli increased activations only in the lateral occipital sulci bilaterally. The opposite effects of sensory reliability in the visual and auditory domain reflect the asymmetry of the visual selective attention paradigm that renders visual information task-relevant and auditory information task-irrelevant.

Accumulation of audiovisual object evidence

Regions that accumulate auditory and visual evidence were expected to exhibit incongruency effects that increase with the variance of the task-relevant visual input and the reliability of the interfering auditory signal (Table 2; supplemental material, available at www.jneurosci.org). The left inferior frontal sulcus extending into the inferior precentral sulcus was the only region that fulfilled these criteria. First, the left inferior frontal sulcus was the only region that showed an interaction between incongruency and visual reliability when correcting for multiple comparisons. As shown in the parameter estimate plots at the peak coordinate of the significant cluster (Fig. 4*A*, right), the incongruency effects emerged primarily when the visual input was unreliable. Second, the left IFS exhibited an interaction between incongruency and auditory reliability. As expected for an accumulator region, the incongruency effects were greater for reliable than unreliable auditory input. This interaction was observed at an uncorrected level of significance. However, as the interactions of incongruency with (1) visual and (2) auditory reliability are orthogonal, the highly significant interaction between visual reliability and incongruency can be used as a search volume constraint for the incongruency × auditory reliability interaction. At an uncorrected level of significance, the right intraparietal sulcus showed a similar pattern as the left IFS, i.e., interactions between incongruency and (1) visual and (2) auditory reliabilities. It is

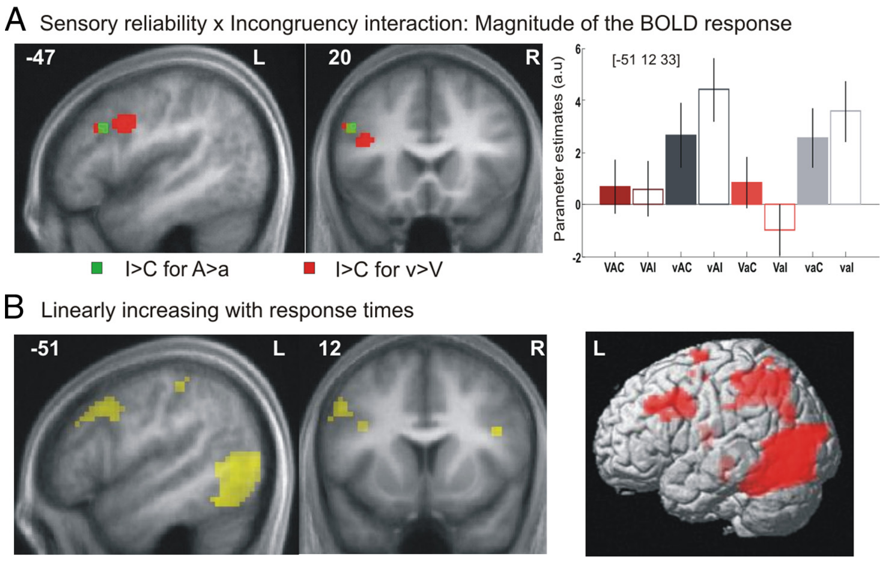
A Sensory reliability x Incongruency interaction: Magnitude of the BOLD response
-47
L
20
R
🟩 I>C for A>a
🟥 I>C for v>V
[-51 12 33]
Parameter estimates (a.u)
6
4
2
0
-2
VAC
VAI
vAC
vAI
VaC
VaI
vaC
vaI
B Linearly increasing with response times
-51
L
12
R
L

Figure 4. fMRI results: accumulation of audiovisual object evidence. ***A***, Left, Sensory reliability × incongruency interactions for the BOLD response magnitude (i.e., canonical hemodynamic response function) on sagittal and coronal slices of a mean structural image created by averaging the subjects' normalized structural images. Red, Incongruency effect increased for degraded relative to intact visual input. Green, Incongruency effect increased for intact relative to degraded auditory input. Height threshold, $p < 0.001$ uncorrected masked with stimulus > fixation at $p < 0.001$. Extent threshold, >5 voxels. Right, Parameter estimates of canonical hemodynamic response function at $x = -51$, $y = 12$, $z = 33$. The bar graphs represent the size of the effect in nondimensional units (corresponding to percentage whole-brain mean). V, Intact vision; v, degraded vision; A, intact audition; a, degraded audition; C, congruent; I, incongruent. ***B***, Activations that are positively predicted by trial-specific response times displayed on sagittal and coronal slices of a mean structural image (left) or rendered on a template of the whole brain (right, only left hemisphere displayed). Height threshold, $p < 0.001$ uncorrected masked with stimulus > fixation at $p < 0.001$. Extent threshold, >0 voxels.

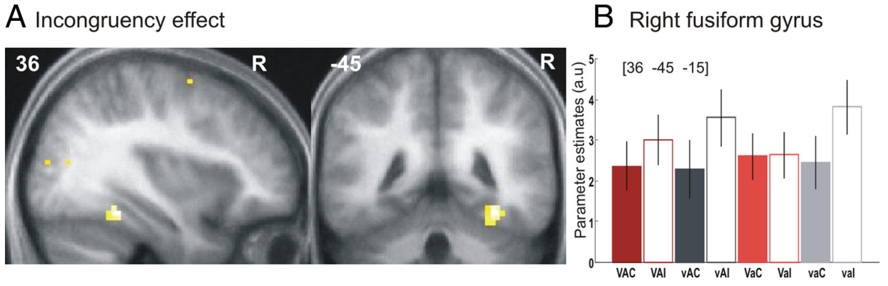

Figure 5. fMRI results: incongruency effects. ***A***, Increased activation for incongruent relative to congruent trials on sagittal and coronal slices of a mean structural image created by averaging the subjects' normalized structural images. Height threshold, $p < 0.001$ uncorrected masked with stimulus > fixation at $p < 0.001$. Extent threshold, > 0 voxels. ***B***, Parameter estimates of canonical hemodynamic response function at $x = 36$, $y = -45$, $z = -15$. The bar graphs represent the size of the effect in nondimensional units (corresponding to percentage whole-brain mean). V, Intact vision; v, degraded vision; A, intact audition; a, degraded audition; C, congruent; I, incongruent.

important to note that the visual and auditory reliabilities modulate the incongruency effects in opposite directions as suggested by the compatibility bias model: Increasing the reliability of the visual input that needs to be categorized reduces the incongruency effect, while increasing the reliability of the interfering auditory input amplifies the incongruency effect. Furthermore, the modulatory effect of visual reliability is greater than that of auditory reliability on audiovisual incongruency in that region (Fig. 4*A*) as a result of visual selective attention. Since the interactions between (in)congruency and sensory reliability were observed in opposing directions for visual and auditory signals, these effects cannot easily be attributed to sensory degradation per se (unlike in some previous unisensory paradigms).

For completeness, we did not observe a significant three-way interaction.

Finally, we used intertrial variability in subjects' response times to identify areas involved in decision making. To this end, we expanded our initial general linear model by one additional regressor that modeled trial-specific reaction times for all trials from all conditions. Note that due to the extra sum of squares principle, the reaction time effects account only for that partition of variance that cannot be explained by the condition effects. The left inferior frontal/precentral sulcus was positively predicted by reaction times ($x = -54$, $y = 15$, $z = 33$, z-score = 4.2; $x = -51$, $y = 21$, $z = 33$, z-score = 3.7). However, as shown in Figure 4*B*, a widespread neural system including bilateral insulae and occipitotemporal and parietal cortices is positively predicted by reaction times after whole-brain correction. Thus, even though reaction times are quite commonly used to identify the neural systems underlying decision making, they may not be sufficiently specific as predictors in the current experimental paradigm, most likely because response times also covary with many other cognitive processes that are unrelated to multisensory evidence accumulation such as stimulus processing times, working memory demands, etc.

Main effect of incongruency

The right fusiform gyrus was the only region showing increased activation for incongruent relative to congruent audiovisual stimuli (Fig. 5). At an uncorrected threshold, a small interaction between incongruency and visual reliability ($x = 36$, $y = -45$, $z = -15$, z-score = 3.0), yet no interaction between incongruency and auditory reliability was observed in this region. This suggests that the right fusiform shows an incongruency effect that is only to a very limited degree modulated by auditory and visual reliabilities. In our previous auditory selective attention paradigm (Noppeney et al., 2008), incongruency effects were observed primarily within the auditory processing system. Similarly, Weisman et al. (2004) have demonstrated a double dissociation of incongruency effects within the visual and auditory cortices depending on whether the visual or auditory signals were task-relevant. Collectively, these results suggest that incongruent signals enhance the neural processes in the task-relevant modality reflecting either amplification of task-relevant information (Miller and Cohen, 2001; Egner and Hirsch, 2005; Miller and D'Esposito, 2005) or audiovisual mismatch or prediction errors (Noppeney et al., 2008).

Consistent with our previous auditory selective attention paradigm, we found activation increases only for incongruent relative to congruent pairs, while no activation increases were observed for congruent relative to incongruent pairs within the neural systems activated relative to baseline. This contrasts with activation results for passive listening and viewing, where congruent audiovisual stimuli that allow successful binding of sensory inputs are associated with increased activation relative to incongruent or unimodal stimuli (Calvert et al., 2000; van Atteveldt et al., 2004; Naumer et al., 2009). These opposite activation patterns highlight the role of task context and attention on the neural processes underlying multisensory integration: activation increases for congruent relative to incongruent stimuli are observed primarily when both auditory and visual signals are attended, relevant, and integrated into a unified percept. Thus, when subjects perform a congruency judgment that requires access and comparison of the two independent unisensory percepts and hence precludes natural audiovisual integration, differences between congruent and incongruent stimulus pairs are attenuated (Beauchamp et al., 2004; van Atteveldt et al., 2007) [for review of semantic incongruency manipulations in audiovisual integration, see Doehrmann and Naumer (2008)].

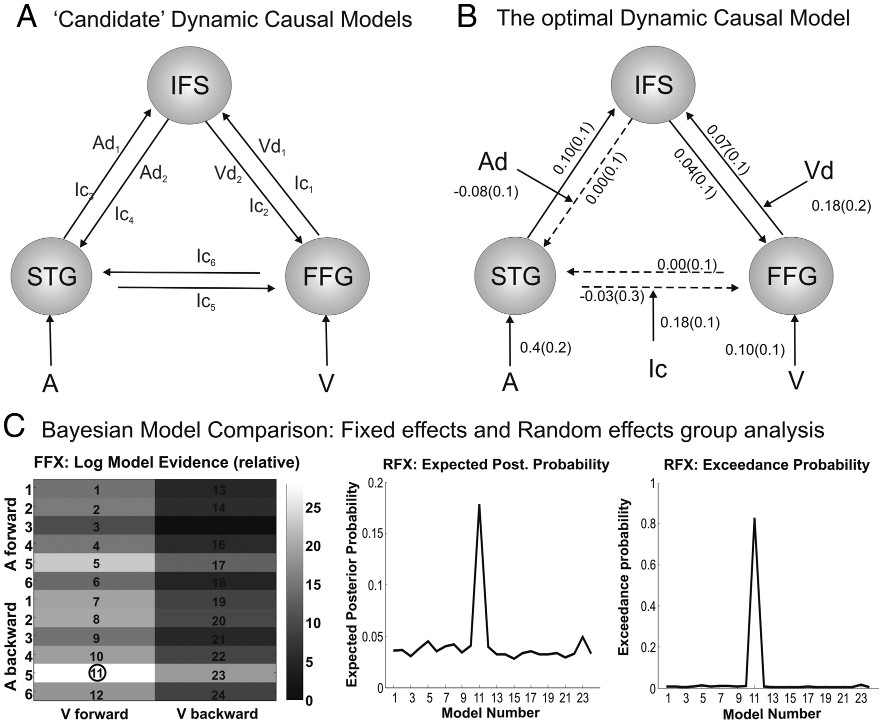

Figure 6. Dynamic causal modeling. ***A***, Twenty-four candidate dynamic causal models were generated by factorially manipulating the connection that was modulated by (1) visual reliability (Vd, visual degraded modulated forward vs backward connection between FFG and IFS), (2) auditory reliability (Ad, auditory degraded modulated forward vs backward connection between STG and IFS), (3) incongruency (Ic, incongruency modulated any of the 6 connections). IFS, Left inferior frontal sulcus. STG, Left superior temporal gyrus. FFG, Right fusiform gyrus. A, All auditory input. V, All visual input. ***B***, In the optimal (highest model evidence) DCM, visual degradation (Vd) modulates the forward and auditory degradation (Ad) the backward connection; the incongruency (Ic) effect enables the connection from STG to FFG. Values are the across-subjects mean (SD) of changes in connection strength ($p < 0.05$ indicated by solid lines). The bilinear parameters quantify how experimental manipulations change the values of intrinsic connections. ***C***, Bayesian model comparison—fixed-effects and random-effects group analysis. Left, Fixed-effects group analysis. The matrix image shows the log model evidence (summed over subjects) of the 24 DCMs in a factorial fashion (model number 1–24). The abscissa shows effects of visual reliability (forward vs backward). The ordinate represents effect of auditory reliability (forward vs backward) × incongruency (each of the 6 connections). The model evidence is higher for models that allow for modulation of the forward connections by visual reliabilities, backward connections by auditory reliabilities, and the connection from STG to FFG by incongruency resulting in the highest model evidence for model 11. Middle/Right, Random-effects group analysis. Expected posterior probability and exceedance probability are shown as a function of model number. Visual reliability modulates the forward connections for model 1–12 and the backwards connections for model 13–24. Consistent with the fixed-effects analysis, model 11 is associated with the highest expected posterior probability of 0.2 and exceedance probability of 0.9.

Surprisingly, the anterior cingulate/medial superior frontal gyrus (AC/mPFC) generally implicated in conflict detection and monitoring (Botvinick et al., 1999; Duncan and Owen, 2000; Botvinick et al., 2001; Paus, 2001; Noppeney and Price, 2002; Laurienti et al., 2003; Kerns et al., 2004; Rushworth et al., 2004; Brown and Braver, 2005; Carter and van Veen, 2007; Orr and Weissman, 2009) showed only a small incongruency effect (coordinates [6 18 42], *z*-score = 2.85; $p = 0.002$ uncorrected). Further, the parameter estimate plots demonstrated that the incongruency effect in the AC/mPFC was most pronounced for visually degraded stimuli. In fact, AC/mPFC was only activated relative to fixation for incongruent visual degraded conditions. Similarly, previous psychophysics studies have shown that incongruent visual signals strongly interfere with auditory object recognition, but incongruent auditory input impedes visual object recognition only when the visual information is rendered less informative (Yuval-Greenberg and Deouell, 2007; Yuval-Greenberg and Deouell, 2009). These findings suggest that vision usually dominates object recognition, rendering it relatively immune to incongruent auditory input. Thus, as previously suggested (Botvinick et al., 2004), AC/mPFC activation may not only signal conflict between multiple sensory inputs, but also the need for cognitive control to override a prepotent incorrect response tendency.

Summary of the results from the conventional (i.e., regional) SPM analysis

In summary, the left inferior frontal sulcus is the only region that fulfilled all three criteria posed for an AV accumulator region: First, the incongruency effect increased with the variance of the visual input. Second, the incongruency effect increased with the reliability of the interfering auditory input. Third, activation in the left IFS increased with increasing response times even when the main effects of condition were modeled in the GLM. In contrast to the left IFS, the right fusiform showed a main effect of incongruency that was only slightly modulated by sensory reliability.

DCM results

Figure 6*C* shows the log evidences summed over subjects for the 24 DCMs relative to the worst model from the fixed-effects group analysis (left). The model evidence is greater for DCMs where (1) visual reliabilities modulate the forward connections from FFG to IFS, (2) auditory reliabilities modulate the backward connections from IFS to STG, and (3) incongruency modulates the connection between STG and FFG. As all 24 DCMs were equated for the number of modulatory effects and intrinsic and extrinsic connectivity structure, the difference in model evidence is only due to model fit and not model complexity. The fixed-effects group analysis provided strong evidence for model 11. Thus, the group Bayes factor was 181 for the optimal model 11 relative to the second best DCM 5.

These results were confirmed in the subsequent random-effects group analysis. As shown in Figure 6*C* (right), model 11 was associated with an expected posterior probability of 0.18, which is very large compared to the posterior probability of 1/24 = 0.04 when assuming a uniform distribution over models. Furthermore, the exceedance probability of model 11 being more likely than any other model tested was 0.83. As the posterior and exceedance probability of a particular model in the random-effects analysis does not only depend on the data but also the set of models tested, we also directly compared the top two models 11 and 23. Again Bayesian model selection revealed a large posterior probability for model 11 of 0.9 and a very high exceedance probability favoring model 11 relative to model 23 of 0.99.

Figure 6*B* shows the changes in connection strength for the optimal model 11. The optimal model 11 included (1) modulation of forward connectivity from FFG to IFS by visual reliability, (2) modulation of backward connectivity from IFS to STG by auditory reliability, and (3) modulation of connectivity from STG to FFG by incongruency. The numbers by the modulatory effects index the change in coupling (i.e., responsiveness of the target region to activity in the source region) induced by visual or auditory reliability or semantic incongruency averaged across subjects. The IFS interacts with the right FFG in an excitatory recurrent loop with the forward connectivity from FFG to IFS being strengthened when the visual input is unreliable. In contrast, the left IFS did not excite and even inhibited activation in STG, when the auditory stimuli were degraded. Furthermore, incongruent audiovisual stimuli enabled the "direct" connection from STG to FFG that was not significantly activated when the auditory and visual inputs were congruent (n.b. "direct" effective connectivity does not imply direct anatomical connectivity). In other words, the right fusiform showed a greater responsiveness to incongruent than congruent auditory input. The optimal DCM highlights three important aspects: First, the visual selective attention or task-relevance was reflected in the asymmetry of the backward connectivity from IFS to auditory and visual areas (Miller and Cohen, 2001; Miller and D'Esposito, 2005). While the IFS exerted an excitatory effect on the FFG, it inhibited activation in the STG. Second, the forward connection from FFG to IFS was enhanced for unreliable visual signals to enhance the accumulation process. Third, incongruent auditory information interfered with visual processing not only in IFS, but already via "direct" effective connectivity between auditory and visual regions (i.e., at higher-order sensory processing levels), suggesting that incongruent auditory inputs may elicit a mismatch or error response in the fusiform area by mechanisms of synaptic integration.

Discussion

Our results suggest that the left IFS accumulates AV object evidence via recurrent loops with auditory and visual cortices that are adjusted according to signal reliability and decisional relevance.

Within the cortical hierarchy, audiovisual interactions have been observed at several levels, including primary sensory (Foxe et al., 2000; Molholm et al., 2002; Schroeder and Foxe, 2002; Kayser et al., 2005; Lakatos et al., 2007; Werner and Noppeney, 2010), higher-order association (Macaluso et al., 2004; Hein et al., 2007; Sadaghiani et al., 2009; Stevenson and James, 2009; Werner and Noppeney, 2009), and prefrontal (Sugihara et al., 2006) [for review, see Calvert and Lewis (2004), Amedi et al. (2005), Schroeder and Foxe (2005), and Ghazanfar and Schroeder (2006)] cortices. This multitude of integration sites may reflect automatic stimulus-driven, perceptual, and decisional audiovisual interactions.

In a visual selective attention paradigm, we introduced AV incongruency that precludes integration of sensory inputs into a coherent precept to focus selectively on AV interactions at the decisional level. Consistent with previous studies (Ben-Artzi and Marks, 1995; Laurienti et al., 2004; Forster and Pavone, 2008; Schneider et al., 2008), conflicting task-irrelevant auditory information interfered with categorization of task-relevant visual object information as indexed by reduced performance accuracy and slower response times for incongruent relative to congruent stimuli. Since both semantically congruent and incongruent signals were presented spatiotemporally coincident, subjects may a priori expect auditory and visual signals to emanate from a common source and hence attempt to integrate the signals regardless of their semantic congruency. Indeed, in support of this compatibility bias hypothesis (Liu et al., 2009; Yu et al., 2009), the decrement in performance accuracy for incongruent trials was particularly pronounced at the onset of a trial that was dominated by subjects' congruency prior; it diminished over the course of a trial when subjects accumulated evidence about the true relationship of the signals. Similarly, as indicated by the response time profile across conditions, the effect of an "inappropriate" congruency prior depended on sensory reliability: the response time difference for incongruent relative to congruent trials was particularly pronounced for degraded visual and intact auditory information.

The framework of evidence accumulation forms a natural link between the time course of behavioral decisions and neuronal activity. In putative "decision" regions, neuronal firing rates have been shown to build up until a decisional threshold is reached and a response is selected (Gold and Shadlen, 2007). This ramp-like neuronal activity has been suggested to reflect accumulation of noisy sensory evidence provided by lower-level sensory areas (Shadlen and Newsome, 2001; Gold and Shadlen, 2002; Ratcliff and Smith, 2004; Sugrue et al., 2004). The present study goes beyond evidence accumulation in a unisensory context [e.g., Binder et al. (2004), Heekeren et al. (2004), and Ho et al. (2009)] and investigates perceptual decisions under more natural conditions, when sensory signals are simultaneously furnished by multiple senses. Consistent with the proposed links between decision making and motor planning, we observed "decision-related" activations in the left IFS, i.e., contralateral to subjects' response hand. In fact, the left IFS was the only region showing incongruency × sensory reliability interactions that were posed as defining features for a multisensory "accumulator" region: The incongruency effects increased for (1) unreliable visual information (to be categorized) and (2) reliable auditory information (task-irrelevant and interfering). The similar profiles for condition-specific BOLD responses and reaction times are consistent with the compatibility bias model that provides a common generative mechanism for both neural and behavioral responses. In line with previous studies that identified decision-related activations by their correlations with reaction times, the activation in the left IFS was also significantly predicted by trial-specific reaction times. Yet, longer reaction times may not only result from extended evidence accumulation but numerous other cognitive processes (e.g., stimulus processing, attention, working memory, etc.); not surprisingly, the reaction time analysis revealed activation beyond IFS in a widespread bilateral neural system encompassing occipitotemporal, parietal, and prefrontal cortices. This lack of specificity limits the role of response times per se in selectively identifying "AV decision-making regions." In contrast, the current analysis prescribes a specific pattern for AV incongruency × sensory reliability interactions that emerge from accumulation of AV object evidence. Admittedly, the dissociation of accumulation from sustained activity (e.g., working memory processes, etc.) is generally hampered by the low temporal resolution of the BOLD response. For instance, the convolution with the hemodynamic response renders the profiles and predictions of (1) blocks of sustained and (2) ramps of accumulation-related neuronal activity nearly indistinguishable, when the durations show only little variation. To dissociate accumulation from sustained activity, a recent study has artificially prolonged the period of evidence accumulation by revealing objects gradually in a stepwise fashion (Ploran et al., 2007). However, protracted categorization over a 10 s period may not be functionally equivalent to rapid natural categorization. Furthermore, gradually revealing an object progressively increases the information provided by the stimulus and hence induces evidence accumulation at the input level, rendering the interpretation of ramp-like cortical activations rather ambiguous. Thus, a future study may need to present movies that dynamically reveal different features of an object while holding the amount of information per frame constant over time.

In contrast to the prefrontal cortex, the right fusiform (FFG) and additional occipitotemporal regions showed increased activations for incongruent relative to congruent trials, even when both visual and auditory information were reliable. Interestingly, audiovisual incongruency effects were observed only along the visual processing stream, while the auditory systems were unaffected. A similar asymmetry was also observed for the effect of sensory reliability: while auditory degradation reduced activation in the auditory system, visual degradation primarily increased activations in the visual system. These asymmetric response profiles in the auditory and visual systems most likely result from the visual selective attention that makes subjects categorize the visual stimuli and ignore the task-irrelevant auditory stimuli. Reduced visual reliability and the presence of an interfering auditory stimulus enhance activations elicited by the task-relevant visual stimulus. In line with the present results, we have previously shown AV incongruency effects only in the auditory system during an auditory selective attention paradigm (Noppeney et al., 2008) (see also Weissman et al., 2004). Similarly, incongruent pairs of face pictures and written proper names have elicited increased activations in the fusiform gyri bilaterally (Egner and Hirsch, 2005). These incongruency effects have been interpreted as a neural mechanism for amplification of task-relevant information to overcome irrelevant and even conflicting auditory information (Egner and Hirsch, 2005). The human brain may resolve conflicts, both within and across the senses, through amplification of task-relevant rather than suppression of incongruent task-irrelevant information. Equally well, however, they may reflect error-related responses indexing a mismatch between visual and auditory inputs (Noppeney et al., 2008).

To investigate how accumulation of AV object evidence and resolution of intersensory conflict emerges from distinct interactions among brain regions, we combined dynamic causal modeling and Bayesian model selection. In the optimal DCM, visual selective attention was reflected in the backward connectivity from IFS to FFG and STG: the IFS exerted an excitatory effect on the fusiform, but an inhibitory effect on the auditory cortex. This excitatory recurrent loop between IFS and FFG was strengthened in its forward connectivity, when accumulation was enhanced for unreliable visual information. Interestingly, incongruent auditory input influenced visual processing not only in IFS but also via "direct" effective connectivity from STG to FFG. Thus, the IFS may integrate evidence directly from visual and auditory regions as well as accumulate integrated audiovisual evidence from FFG. Our DCM results may also lend themselves to an interpretation of evidence accumulation in terms of predictive coding where perceptual inference is implemented in recurrent message passing between multiple levels within the cortical hierarchy (Rao and Ballard, 1999; Summerfield and Koechlin, 2008; Friston, 2009; Friston and Kiebel, 2009). The backwards connections hereby encode the top-down prior predictions; the forward connections furnish the residual errors between the predictions and the actual incoming inputs. During "evidence accumulation," prediction errors at all levels of the cortical hierarchy are used to guide perceptual inference. From this perspective, the IFS learns high-level representations to enable response selection based on prediction errors furnished via forward connectivity from auditory and visual areas. The increase in prediction error for less predictable unreliable visual signals is manifest in increased forward connectivity from FFG to IFS when the visual input is degraded. Similarly, the violation of subjects' prior congruency assumption is manifest in increased STG to FFG connectivity for incongruent auditory input. In other words, incongruent auditory inputs elicit a prediction error response in the right fusiform. In recurrent loops with auditory and visual regions, the IFS may progressively adjust its representations throughout the course of a trial to predict the incoming audiovisual inputs.

In conclusion, to form categorical decisions in our multisensory environment, the IFS may accumulate audiovisual evidence by dynamically weighting its connectivity to auditory and visual regions according to sensory reliability and decisional relevance.

References

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

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

Ben-Artzi E, Marks LE (1995) Visual-auditory interaction in speeded classification: role of stimulus difference. Percept Psychophys 57:1151–1162.

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci 7:295–301.

Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD (1999) Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature 402:179–181.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108:624–652.

Botvinick MM, Cohen JD, Carter CS (2004) Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci 8:539–546.

Brown JW, Braver TS (2005) Learned predictions of error likelihood in the anterior cingulate cortex. Science 307:1118–1121.

Calvert GA, Lewis JW (2004) Hemodynamic studies of audio-visual interactions. In: The handbook of multi-sensory processes (Calvert GA, Spence C, Stein BE, eds), pp 483–502. Cambridge: MIT Press.

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.

Carter CS, van Veen V (2007) Anterior cingulate cortex and conflict detection: an update of theory and data. Cogn Affect Behav Neurosci 7:367–379.

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.

Dakin SC, Hess RF, Ledgeway T, Achtman RL (2002) What causes nonmonotonic tuning of fMRI response to noisy images? Curr Biol 12:R476 –R477.

de Lange FP, Jensen O, Dehaene S (2010) Accumulation of evidence during sequential decision making: the importance of top-down factors. J Neurosci 30:731–738.

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

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

Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends Neurosci 23:475– 483.

Egner T, Hirsch J (2005) Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. Nat Neurosci 8:1784 –1790.

Evans AC, Collins DL, Milner B (1992) An MRI-based stereotactic atlas from 250 young normal subjects. Soc Neurosci Abstr 18:408.

Forster B, Pavone EF (2008) Electrophysiological correlates of crossmodal visual distractor congruency effects: evidence for response conflict. Cogn Affect Behav Neurosci 8:65–73.

Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Brain Res Cogn Brain Res 10:77– 83.

Friston K (2009) The free-energy principle: a rough guide to the brain? Trends Cogn Sci 13:293–301.

Friston K, Kiebel S (2009) Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci 364:1211–1221.

Friston KJ, Holmes A, Worsley KJ, Poline JB, Frith CD, Frackowiak R (1995) Statistical parametric mapping: a general linear approach. Hum Brain Mapp 2:189 –210.

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

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

Gold JI, Shadlen MN (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron 36:299 –308.

Gold JI, Shadlen MN (2007) The neural basis of decision making. Annu Rev Neurosci 30:535–574.

Grinband J, Hirsch J, Ferrera VP (2006) A neural representation of categorization uncertainty in the human brain. Neuron 49:757–763.

Heekeren HR, Marrett S, Bandettini PA, Ungerleider LG (2004) A general mechanism for perceptual decision-making in the human brain. Nature 431:859 – 862.

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.

Ho TC, Brown S, Serences JT (2009) Domain general mechanisms of perceptual decision making in human cortex. J Neurosci 29:8675– 8687.

Kaiser J, Lennert T, Lutzenberger W (2007) Dynamics of oscillatory activity during auditory decision making. Cereb Cortex 17:2258 –2267.

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90:773–795.

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

Kerns JG, Cohen JD, MacDonald AW 3rd, Cho RY, Stenger VA, Carter CS (2004) Anterior cingulate conflict monitoring and adjustments in control. Science 303:1023–1026.

Kiani R, Hanks TD, Shadlen MN (2008) Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. J Neurosci 28:3017–3029.

Kim JN, Shadlen MN (1999) Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neurosci 2:176 –185.

Kleiner M, Wallraven C, Bulthoff HH (2004) The MPI VideoLab—a system for high quality synchronous recording of video and audio from multiple viewpoints. Tübingen, Germany: MPI 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.

Laurienti PJ, Wallace MT, Maldjian JA, Susi CM, Stein BE, Burdette JH (2003) Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. Hum Brain Mapp 19:213–223.

Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT (2004) Semantic congruence is a critical factor in multisensory behavioral performance. Exp Brain Res 158:405– 414.

Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA (2004) Human brain regions involved in recognizing environmental sounds. Cereb Cortex 14:1008 –1021.

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.

Liu YS, Yu A, Holmes P (2009) Dynamical analysis of Bayesian inference models for the Eriksen task. Neural Comput 21:1520 –1553.

Lo CC, Wang XJ (2006) Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nat Neurosci 9:956 –963.

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

Mazurek ME, Roitman JD, Ditterich J, Shadlen MN (2003) A role for neural integrators in perceptual decision making. Cereb Cortex 13:1257–1269.

Miller BT, D'Esposito M (2005) Searching for "the top" in top-down control. Neuron 48:535–538.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.

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. Brain Res Cogn Brain Res 14:115–128.

Naumer MJ, Doehrmann O, Müller NG, Muckli L, Kaiser J, Hein G (2009) Cortical plasticity of audio-visual object representations. Cereb Cortex 19:1641–1653.

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

Orr JM, Weissman DH (2009) Anterior cingulate cortex makes 2 contributions to minimizing distraction. Cereb Cortex 19:703–711.

Paus T (2001) Primate anterior cingulate cortex: where motor control, drive and cognition interface. Nat Rev Neurosci 2:417– 424.

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

Philiastides MG, Sajda P (2006) Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex 16:509 –518.

Pleger B, Ruff CC, Blankenburg F, Bestmann S, Wiech K, Stephan KE, Capilla A, Friston KJ, Dolan RJ (2006) Neural coding of tactile decisions in the human prefrontal cortex. J Neurosci 26:12596 –12601.

Ploran EJ, Nelson SM, Velanova K, Donaldson DI, Petersen SE, Wheeler ME (2007) Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. J Neurosci 27:11912–11924.

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79 – 87.

Ratcliff R, Smith PL (2004) A comparison of sequential sampling models for two-choice reaction time. Psychol Rev 111:333–367.

Romo R, Hernández A, Zainos A (2004) Neuronal correlates of a perceptual decision in ventral premotor cortex. Neuron 41:165–173.

Rushworth MF, Walton ME, Kennerley SW, Bannerman DM (2004) Action sets and decisions in the medial frontal cortex. Trends Cogn Sci 8:410 – 417.

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.

Schall JD (2003) Neural correlates of decision processes: neural and mental chronometry. Curr Opin Neurobiol 13:182–186.

Schneider TR, Debener S, Oostenveld R, Engel AK (2008) Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. Neuroimage 42:1244–1254.

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. Brain Res Cogn Brain Res 14:187–198.

Servan-Schreiber D, Bruno RM, Carter CS, Cohen JD (1998a) Dopamine and the mechanisms of cognition: Part I. A neural network model predicting dopamine effects on selective attention. Biol Psychiatry 43:713–722.

Servan-Schreiber D, Carter CS, Bruno RM, Cohen JD (1998b) Dopamine and the mechanisms of cognition: Part II. D-amphetamine effects in human subjects performing a selective attention task. Biol Psychiatry 43:723–729.

Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86:1916–1936.

Smith PL, Ratcliff R (2004) Psychology and neurobiology of simple decisions. Trends Neurosci 27:161–168.

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.

Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. Science 304:1782–1787.

Summerfield C, Koechlin E (2008) A neural representation of prior information during perceptual inference. Neuron 59:336–347.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme.

Thielscher A, Pessoa L (2007) Neural correlates of perceptual choice and decision making during fear-disgust discrimination. J Neurosci 27:2908–2917.

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 NM, Formisano E, Goebel R, Blomert L (2007) Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. Neuroimage 36:1345–1360.

Weissman DH, Warner LM, Woldorff MG (2004) The neural mechanisms for minimizing cross-modal distraction. J Neurosci 24:10941–10949.

Werner S, Noppeney U (2009) Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cereb Cortex. Advance online publication. Retrieved May 17, 2010. doi:10.1093/cercor/bhp248.

Werner S, Noppeney U (2010) Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J Neurosci 30:2662–2675.

Yu AJ, Dayan P, Cohen JD (2009) Dynamics of attentional selection under conflict: toward a rational Bayesian account. J Exp Psychol Hum Percept Perform 35:700–717.

Yuval-Greenberg S, Deouell LY (2007) What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. J Neurosci 27:1090–1096.

Yuval-Greenberg S, Deouell LY (2009) The dog's meow: asymmetrical interaction in cross-modal object recognition. Exp Brain Res 193:603–614.