Superadditive Responses in Superior Temporal Sulcus Predict Audiovisual Benefits in Object Categorization

Sebastian Werner and Uta Noppeney

Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

Merging information from multiple senses provides a more reliable percept of our environment. Yet, little is known about where and how various sensory features are combined within the cortical hierarchy. Combining functional magnetic resonance imaging and psychophysics, we investigated the neural mechanisms underlying integration of audiovisual object features. Subjects categorized or passively perceived audiovisual object stimuli with the informativeness (i.e., degradation) of the auditory and visual modalities being manipulated factorially. Controlling for low-level integration processes, we show higher level audiovisual integration selectively in the superior temporal sulci (STS) bilaterally. The multisensory interactions were primarily subadditive and even suppressive for intact stimuli but turned into additive effects for degraded stimuli. Consistent with the inverse effectiveness principle, auditory and visual informativeness determine the profile of audiovisual integration in STS similarly to the influence of physical stimulus intensity in the superior colliculus. Importantly, when holding stimulus degradation constant, subjects' audiovisual behavioral benefit predicts their multisensory integration profile in STS: only subjects that benefit from multisensory integration exhibit superadditive interactions, while those that do not benefit show suppressive interactions. In conclusion, superadditive and subadditive integration profiles in STS are functionally relevant and related to behavioral indices of multisensory integration with superadditive interactions mediating successful audiovisual object categorization.

Keywords: crossmodal, fMRI, inverse effectiveness, multisensory, object recognition

Introduction

To interact effectively with our multisensory environment, the human brain integrates information from multiple sources into a coherent percept. Multisensory integration facilitates detection, identification, and categorization of objects and events. Neurophysiological and functional imaging studies in human and nonhuman primates have revealed multisensory interactions in a widespread system encompassing subcortical structures (Wallace et al. 1996; Calvert et al. 2001), putative unisensory cortices (Schroeder and Foxe 2002; Molholm et al. 2004; van Atteveldt et al. 2004; Ghazanfar et al. 2005; Bonath et al. 2007; Kayser et al. 2007; Lakatos et al. 2007; Martuzzi et al. 2007), and higher order association cortices (Macaluso et al. 2003; Beauchamp, Argall, et al. 2004; Barraclough et al. 2005; Miller and D'Esposito 2005; Noesselt et al. 2007; Ghazanfar et al. 2008; Noppeney et al. 2008; Sadaghiani et al. 2009). However, the types of information that are integrated in this multitude of integration sites remain unclear: spatial (where?), temporal (when?), object-related (what?), and other types of information may be integrated at different levels of the cortical hierarchy.

The neural mechanisms responsible for integrating these various stimulus parameters are poorly understood. Neurophysiological studies have traditionally focused on enhanced responses for 2 stimuli from different modalities relative to the most effective unisensory stimulus (Stein and Meredith 1993). This multisensory enhancement (MSE) can range from subadditive to additive or even superadditive combinations of modality-specific inputs (Perrault et al. 2005; Stanford et al. 2005). According to the inverse effectiveness principle, MSE is maximal when the individual inputs are least effective (Meredith and Stein 1983; Stanford and Stein 2007). In addition to these excitatory–excitatory multisensory interactions (i.e., both sensory inputs are excitatory), more recently, an excitatory–inhibitory form has been reported where bisensory responses are less than the most effective unisensory response (i.e., one sensory input is excitatory and the other inhibitory) (Meredith 2002). Importantly, this suppressive (and thus subadditive) form of multisensory integration has been observed even in the absence of spatiotemporal misalignments (Dehner et al. 2004; Barraclough et al. 2005; Sugihara et al. 2006; Avillac et al. 2007) raising questions about the functional relevance of the diverse (i.e., subadditive vs. superadditive) multisensory integration profiles.

The aim of this study was 2-fold: First, the experiment was designed to isolate integration of higher order auditory and visual object features and the underlying neural processes during categorization. Second, we characterized the computations underlying audiovisual integration of higher order features. In particular, we examined the factors that determine whether audiovisual information is integrated in a superadditive, additive, or subadditive fashion.

We manipulated the informativeness (intact, degraded, and noise) of concurrently presented auditory and visual stimuli in a $3 \times 3 \times 2$ factorial design (Fig. 1). Subjects actively categorized audiovisual stimuli as tools or musical instruments or passively attended to them in a target detection task. All experimental conditions were equated with respect to low-level stimulus characteristics (i.e., spatial/temporal frequency contents and low-level image/sound statistics) but differed in the availability of higher order object-related information.

We then addressed the following 3 issues: First, we identified the neural systems that integrate higher level information about an object's category while controlling for audiovisual integration of low-level spatiotemporal stimulus features. Second, we formally investigated the inverse effectiveness principle and characterized underlying computations as subadditive and superadditive at multiple levels of stimulus informativeness.
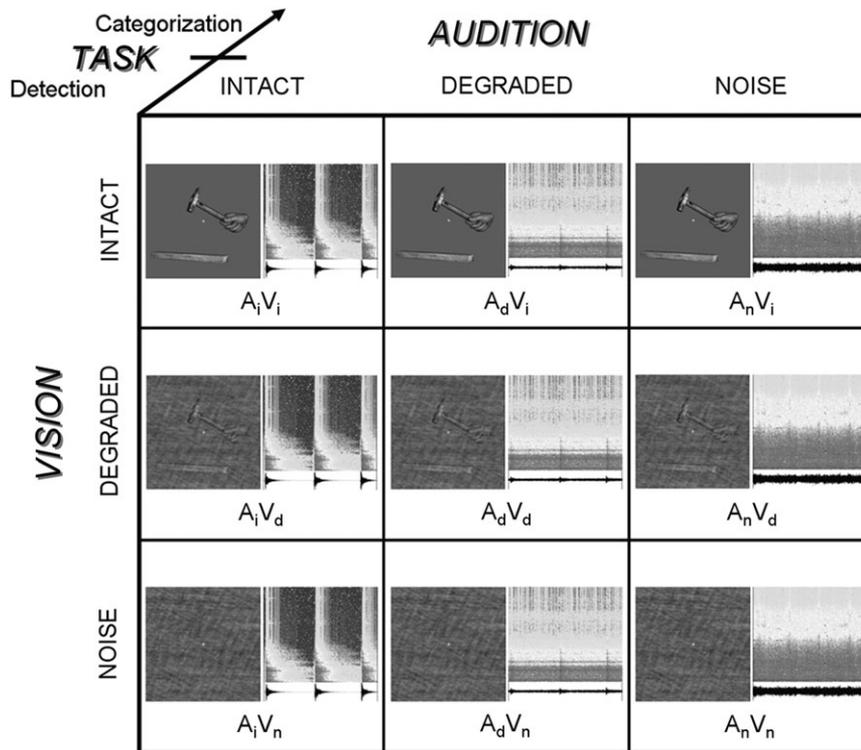
Figure 1. Experimental design and example stimuli. A 3 × 3 × 2 factorial design with the factors: 1) Visual informativeness: intact, degraded, and noise; 2) Auditory informativeness: intact, degraded, and noise; 3) Task: categorization and target detection. Example stimuli are presented as visual images and corresponding sound spectrograms (0–2.5 kHz, 2 s) with amplitude waveforms. Object-related information was manipulated by applying different levels of Fourier phase scrambling that preserve low-level stimulus characteristics (i.e., spatial/temporal frequency components and low-level image/sound statistics) across stimulus conditions. A_i = Auditory intact; A_d = Auditory degraded; A_n = Auditory noise; V_i = Visual intact; V_d = Visual degraded; V_n = Visual noise;

Third, we evaluated the functional relevance of the audiovisual integration profiles by relating them to behavioral indices of subject's multisensory benefits.

Materials and Methods

Subjects

Twenty right-handed subjects (10 females; mean age: 25.8 years; standard deviation: 4.5) 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. Handedness was determined based on self-report. The study was approved by the Human Research Ethics Committee of the Medical Faculty at the University of Tübingen.

Stimulus Presentation

Visual and auditory stimuli were presented using Cogent (John Romaya, Vision Lab, University College London, UK; http://www.vislab.ucl.a-c.uk/), running under Matlab 7.0 (MathWorks Inc., Natick, MA) on a Windows PC. Auditory stimuli were presented at approximately 81 dB SPL, using magnetic resonance (MR)-compatible headphones (MR Confon GmbH, Magdeburg, Germany). Visual stimuli (size 6.5° × 6.5° visual angle) 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. Subjects performed a behavioral task using a MR-compatible custom-built button device connected to the stimulus computer.

Stimuli

Visual stimuli were gray-scale photographs of 15 tools (e.g., hammer, saw, drill, and scissors) and 15 musical instruments (e.g., drum, guitar, flute, and violin) taken from video clips recorded at the Max Planck Institute VideoLab (Kleiner et al. 2004). The 2 distinct categories were selected to allow for a semantic categorization task. Yet, category-

selective activations are not the focus of this communication (Chao et al. 1999; Lewis et al. 2004, 2005; Noppeney et al. 2006). Informativeness of the images was manipulated by applying different degrees of Fourier phase scrambling. To this end, original (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. Three levels of informativeness were generated by combining the original amplitude spectra with 1) the original phase components (i.e., "intact vision"), 2) phase components representing a linear interpolation between original and random noise phase spectra (i.e., "degraded vision"), or 3) the phase components of uniform random noise images (i.e., "noise vision"). The linear interpolation between the original and random noise phase spectra (i.e., degraded vision) preserved 20% of the original phase components to enable threshold performance (~75% accuracy) based on behavioral pilot studies (note: in the functional imaging study, % accuracy was reduced to 66% only). The procedure ensured that individual images at different levels of informativeness 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 images from the 2 categories (i.e., tools or musical instruments) with respect to their mean luminance (t_{28} = 1.39; P > 0.05) and RMS contrast (t_{28} = 0.16; P > 0.05).

Auditory stimuli were sounds produced by actions of the tools and musical instruments that were used as visual stimuli (see above). Each sound clip (2-s duration, 48-kHz sampling rate) was presented monophonically and was equated for maximum intensity of the sound stimulation. Similar to the visual domain, original and white noise sounds were transformed into Fourier amplitude and phase components. Three levels of informativeness were generated by combining the original temporal frequency amplitude spectra with 1) the original phase components (i.e., "intact sound"), 2) phase components representing a linear interpolation between original and white noise

phase spectra (i.e., "degraded audition"), or 3) by combining the average of all original frequency amplitude spectra (i.e., of all tools and musical instruments) with the phase components of auditory white noise (i.e., "noise audition"). The linear interpolation between the original and white noise phase spectra (i.e., degraded audition) preserved 30% of the original phase components to enable threshold performance (~75% accuracy) based on behavioral pilot studies (note: ~75% accuracy was also obtained in the subsequent functional magnetic resonance imaging [fMRI] study). The procedure ensured that sounds across different levels of informativeness were matched in terms of their temporal frequency contents, distribution of phase components, and RMS power. To prevent subjects from using low-level auditory cues for categorization, we selected and matched the sounds from the 2 categories (i.e., tools or musical instruments) with respect to their RMS power ($t_{28} = 1.12$; $P > 0.05$).

In addition to the object stimuli, simple stimuli (i.e., a circle, a tone, or a circle + tone) were included as targets during those periods when subjects were engaged in a target detection task.

Experimental Design

The currently most stringent methodological approach identifies multisensory integration (Calvert and Lewis 2004) through the interaction between presence and absence of information from 2 modalities (i.e., (AV – rest) – ([A – rest] + [V – rest])), which is equivalent to the superadditivity criterion in neurophysiological studies, that is, the bisensory response exceeds the sum of the unisensory responses after adjustment for baseline activity (Perrault et al. 2005; Sugihara et al. 2006). Similarly, subadditive (and even suppressive) effects can be identified as negative interactions. However, this classical interaction design gives rise to 3 major drawbacks. First, by definition, the interaction term can only identify nonlinear combinations of modality-specific inputs, leaving out additive multisensory integration effects that have been observed at the single neuron level (Perrault et al. 2005; Stanford et al. 2005). Second, for the interaction term to be valid and unbiased, the use of "rest" (= absence of auditory and visual information) precludes that subjects perform a task on the stimuli (Beauchamp, Lee, et al. 2004; Beauchamp 2005). This is because task-related activations are absent during the rest condition leading to an overestimation of the summed unisensory relative to the bisensory fMRI responses in the interaction term. Third, during recognition of complex environmental stimuli such as speech, objects, or actions, multisensory interactions can emerge at multiple processing levels, ranging from integration of low-level spatiotemporal to higher level object-related perceptual information. These different types of integration processes are all included in the statistical comparison (i.e., interaction) when using a rest condition. Hence, a selective dissociation between integration processes of spatiotemporal and object-related information is not possible (for further methodological considerations, see Supplementary Material).

Therefore, we have transformed the classical interaction design into a more elaborate 3 × 3 × 2 factorial design manipulating 1) visual informativeness (intact = V_i, degraded = V_d, and noise = V_n), 2) auditory informativeness (intact = A_i, degraded = A_d, and noise = A_n), and 3) task (categorization, target detection). Analogous to the classical interaction approach, we will use the following stimulus nomenclature. A "unisensory visual" stimulus refers to a visual object image combined with auditory noise. Similarly, "unisensory auditory" refers to auditory object sounds presented with visual noise. In other words, a "unisensory" stimulus is a shorthand for saying that object information was limited to one modality only. Visual and auditory informativeness were manipulated using the Fourier phase scrambling technique that gradually degraded information about object identity while preserving low-level stimulus characteristics. Subjects either actively categorized the audiovisual stimuli as tools or musical instruments or passively attended to these stimuli while engaged in a target detection task. During the target detection task, subjects had to detect and respond to simple target stimuli (i.e., circle, tone, or circle + tone), while they only passively viewed the audiovisual object stimuli. The target detection task enabled us to characterize audiovisual integration processes in the absence of explicit responses.

This experimental design served 3 purposes. First, equating conditions with respect to low-level stimulus characteristics and thus controlling for low-level audiovisual interactions allowed us to focus on higher level multisensory integration of the behaviorally relevant audiovisual object features. Second, independently manipulating the informativeness of the auditory and visual modalities as a surrogate of stimulus efficacy enabled us to investigate the principle of inverse effectiveness. Third, using both a categorization and a target detection task enabled us to distinguish between automatic (target detection) and task-induced (categorization) multisensory interactions.

Experimental Procedure

Subjects performed 2 functional imaging sessions (~17 min each). Each stimulus was presented for 2000 ms followed by 800-ms fixation. Each of the 30 items (15 tools and 15 musical instruments) was presented once in each of the 18 conditions. During the categorization task, subjects categorized audiovisual stimuli as tools or musical instruments. During the target detection task, they responded to simple visual (circle), auditory (tone), or audiovisual (circle + tone) targets, while passively attending to the object stimuli (i.e., tools and musical instruments). Approximately 15% of total events during the target detection task were targets. They were presented for 300 ms (followed by 800-ms fixation). Subjects indicated their response as quickly and accurately as possible via a 2-choice keypress in the categorization task or a single keypress in the target detection task. 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 (stimulus onset asynchrony = 2800 ms) were presented in blocks of 8 stimuli (pseudorandomly selected from the different conditions) interleaved with 6-s fixation. During the detection task, each block contained at least one target (40%: 2 targets). A pseudorandomized stimulus sequence was generated for each subject. The order of task conditions was counterbalanced within and across subjects. In summary, we used a mixed block/event-related design: 1) the task was manipulated in long periods (i.e., a continuous half of each scanning session was dedicated to a single task), 2) stimuli blocks (~22.4 s) alternated with short fixation blocks (6 s), 3) within the stimulus blocks, events of the 9 conditions were presented in a pseudorandomized fashion.

Behavioral Measurements

Subjects' performance measures (% correct, median reaction time) were entered into a repeated measurement analysis of variance (ANOVA). Because of the apparent near-ceiling performance during the intact conditions, signal sensitivity measure d' [= $Z(P_{hits})$ – $Z(P_{false\ alarms})$] was computed only for the degraded conditions, that is, the unisensory visual A_nV_d, unisensory auditory A_dV_n, and bisensory A_dV_d object stimuli. Multisensory benefits were computed as the difference between the bisensory d' and the best, that is, greatest unisensory d' and used to form regressors in a second-level regression model to predict fMRI activations. This multisensory benefit was also computed only for the degraded conditions, as the intact conditions were associated with near-ceiling performance.

To investigate whether subjects efficiently integrated audiovisual information during object categorization, we compared the empirical d' for the bisensory trials to the d' predicted by the probability summation model (PSM) based on the 2 unisensory conditions. The prediction of the PSM is calculated from the relevant 2 unisensory d' 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 PSM is formed from 2 lines (i.e., the 2 unisensory decision bounds) at right angle. We have applied this model to our 2 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 probabilities of hits and false alarms were computed from the unisensory degraded conditions as follows:

Probablity of hit

$$P_{\text{hits}}(AV) = P_{\text{hits}}(A) + P_{\text{hits}}(V) - P_{\text{hits}}(A) \times P_{\text{hits}}(V)$$

Probability of false alarm

$$P_{\text{false alarms}}(AV) = P_{\text{false alarms}}(A) + P_{\text{false alarms}}(V) - P_{\text{false alarms}}(A) \times P_{\text{false alarms}}(V)$$

An empirical d' that is significantly greater than predicted by the probability summation suggests that subjects have not independently processed but integrated the information from the 2 input modalities to some extent (Treisman 1998).

MRI

A 3-T Siemens Magnetom Trio system (Siemens, Erlangen, Germany) was used to acquire both T_1-weighted anatomical volume images and T_2^*-weighted axial echo-planar images with blood oxygen level–dependent (BOLD) contrast (gradient-echo echo-planar imaging, Cartesian k-space sampling, time repetition [TR] = 3080 ms, time echo [TE] = 40 ms, flip angle = 90°, field of view [FOV] = 192 × 192 mm, image matrix = 64 × 64, 38 slices acquired sequentially in ascending direction, 3.0 × 3.0 × 2.6 mm voxels, interslice gap = 0.4 mm). There were 2 sessions with a total of 343 volume images per session. The first 5 volumes were discarded to allow for T_1 equilibration effects. A 3D high-resolution anatomical image was acquired (TR = 10.55 ms, TE = 3.14 ms, time to inversion = 680 ms, flip angle = 22°, FOV = 256 × 224 × 176 mm, image matrix = 256 × 224 × 176, isotropic spatial resolution = 1 mm).

Data Analysis

The data were analyzed with statistical parametric mapping (using SPM2 software from the Wellcome Department of Imaging Neuroscience, UK; 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 Montreal Neurological Institute (MNI) standard space (Evans et al. 1992), resampled to 3 × 3 × 3 mm voxels, and spatially smoothed with a 3D Gaussian kernel of 8-mm full width at half maximum. The time series in each voxel were 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 (representing a single trial) with a canonical hemodynamic response function and its first temporal derivative. In addition to modeling the 18 conditions in our 3 × 3 × 2 factorial design (correct trials only), the first statistical model included targets, instructions, and error trials. Nuisance covariates included the realignment parameters (to account for residual motion artifacts). In a second supplementary analysis, reaction times were modeled for all trials from all conditions as one single regressor to account for differences in processing time across conditions (see Supplementary Table 2). Condition-specific effects for each subject were estimated according to the general linear model and passed to a second-level analysis as contrasts (including only the canonical hemodynamic response function). This involved creating the following contrast images (pooled, i.e., averaged over tasks and sessions):

(1) Superadditivity for intact stimuli:
$(A_iV_i - A_nV_n) - ([A_nV_i - A_nV_n] + [A_iV_n - A_nV_n])$
$= (A_iV_i + A_nV_n) - (A_nV_i + A_iV_n)$.

(2) Superadditivity for degraded stimuli:
$(A_dV_d - A_nV_n) - ([A_nV_d - A_nV_n] + [A_dV_n - A_nV_n])$
$= (A_dV_d + A_nV_n) - (A_nV_d + A_dV_n)$.

(3) Subadditivity for intact stimuli:
$([A_nV_i - A_nV_n] + [A_iV_n - A_nV_n]) - (A_iV_i - A_nV_n)$
$= (A_nV_i + A_iV_n) - (A_nV_n + A_iV_i)$.

(4) Subadditivity for degraded stimuli:
$([A_nV_d - A_nV_n] + [A_dV_n - A_nV_n]) - (A_dV_d - A_nV_n)$
$= (A_nV_d + A_dV_n) - (A_nV_n + A_dV_d)$.

(5) Inverse effectiveness (superadditivity$_{\text{degraded}}$ > superadditivity$_{\text{intact}}$) based on stimulus informativeness:
$[(A_dV_d + A_nV_n) - (A_nV_d + A_dV_n)] - [(A_iV_i + A_nV_n) - (A_nV_i + A_iV_n)]$
$= (A_dV_d + A_nV_i + A_iV_n) - (A_iV_i + A_nV_d + A_dV_n)$.

These contrast images were entered into independent second-level 1-sample t-tests. In addition, contrast (2) images (i.e., superadditive interaction contrast for degraded stimuli) were entered into a second-level regression analysis that used the subjects' multisensory perceptual benefit (see, Behavioral Measurements) as a predictor for superadditive BOLD–responses. Inferences were made at the second level to allow for a random effects analysis and inferences at the population level (Friston et al. 1999).

Contrasts (1) to (4) define higher level multisensory integration in terms of interactions between levels of visual and auditory informativeness. In other words, we tested whether the activation due to a change in informativeness of the visual input depends on the informativeness of the auditory input. This enables us to focus selectively on the neural processes underlying integration of object information rather than low-level spatiotemporal integration. The interaction term can also be rewritten such that it relates more closely to the classical superadditivity criterion (Calvert and Lewis 2004) that is corrected for baseline activity during the rest condition: (AV – rest) – ([A – rest] + [V – rest]). However, in our design, the rest condition was replaced by a low-level audiovisual noise condition (i.e., A_nV_n) that permits controlling for integration processes of low-level spatiotemporal audiovisual features. Furthermore, it enabled us to test for superadditive and subadditive interactions during the categorization task in an unbiased fashion. This was because subjects also performed the categorization task on low-level audiovisual noise trials. Thus, response selection and other task-related processes were present in all conditions. In this way, the more elaborate factorial design rendered the interaction (AV + noise) $\neq$ (A + V) more balanced with respect to task-induced processes. In fact, in contrast to the traditional interaction (AV + rest) $\neq$ (A + V), task-induced processing might have been enhanced for the audiovisual noise trials (i.e., forced guesses), as indicated by the long reaction times.

Three approaches were used to investigate and characterize the "principle of inverse effectiveness." First, we used stimulus manipulations, that is, the sensory informativeness. Within subjects, we compared audiovisual interactions for degraded and intact stimuli. Thus, we investigated whether the superadditivity is greater for degraded stimuli than intact stimuli or conversely whether the subadditivity is greater for intact stimuli than degraded stimuli (for similar approach, see also Stevenson et al. 2009). Note that under the null hypothesis no relationship would be expected between stimulus informativeness and the pattern of audiovisual interactions at the neural level. Hence, testing for inverse effectiveness based on stimulus manipulations is unbiased and statistically valid. Second, we characterized the principle of inverse effectiveness with the help of the intersubject variability in performance accuracy, that is, the multisensory benefit. More specifically, we used subjects' multisensory benefit as an explanatory variable in a second-level regression analysis to predict superadditive BOLD–responses. Under the null hypothesis, the subjects' multisensory benefit is assumed to be independent from their multisensory neural responses rendering this approach statistically valid. While the first within-subject approach is similar to standard approaches employed in classical neurophysiology, the second approach is based on intersubject variability and thus deviates from classical neurophysiological analysis approaches. Third, in line with previous studies (Perrault et al. 2003, 2005; Kayser et al. 2008), we characterized the inverse effectiveness principle based on endogenous intersubject variability in the regional responsiveness to unisensory inputs. This third approach can be problematic because of inherent statistical dependencies between unisensory responses and the interaction between bisensory und unisensory responses. For the interested reader, we have included analysis, results, and discussion of this third approach in the Supplementary Materials. Please note that strictly speaking only the third approach tests the principle of inverse effectiveness by formally relating superadditivity to regional responsiveness of unisensory inputs. In contrast, the first approach uses "stimulus informativeness" as a surrogate for stimulus efficacy. The second approach further characterizes the relationship between superadditivity and multisensory benefit.

To increase estimation efficiency, all contrasts were pooled across tasks. For descriptive purposes, we also report the statistics at the peak

voxels identified via the pooled contrasts separately for each task. Please note that the task-selective contrasts are not statistically independent from the pooled contrasts. Nevertheless, they provide additional information and enable a more thorough evaluation of the observed activation pattern. For instance, multisensory integration effects that are also observed for the target detection task alone suggest that they emerge at the level of stimulus and perceptual processing and do not simply reflect executive processing (e.g., responses selection).

Search Volume

Each effect was tested for in 2 search volumes. The first search volume included all voxels, that is, the whole brain (54 518 voxels). In order to increase the sensitivity of the analysis with respect to the superior temporal sulci (STS) as an a priori candidate region for audiovisual object recognition (Amedi et al. 2005), the second search volume was limited to the subset of 3764 voxels that were located within Heschl's gyrus (HG), middle temporal gyrus (MTG), and superior temporal gyri (STG) bilaterally as defined by the automated anatomical labeling 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 the search volume (i.e., the entire brain or the STS) using an auxiliary uncorrected voxel threshold of $P < 0.001$ (Friston et al. 1994). Results of the random effects analysis were superimposed onto a T_1-weighted averaged normalized brain of all participants of the study, using the MRIcro software (http://www.sph.sc.edu/comd/rorden/mricro.html).

Results

Behavioral Data

Categorization Task

Subjects categorized the audiovisual stimuli as tools or musical instruments. For performance accuracy, a 2-way repeated measurement (2-way RM) ANOVA with the factors vision (intact, degraded, and noise) and audition (intact, degraded, and noise) identified significant main effects of visual ($F_{1.9,36.2} = 145.3$; $P < 0.001$) and auditory information ($F_{1.4,26.7} = 137.1$; $P < 0.001$) after Greenhouse–Geisser correction. In addition, there was a significant interaction effect between visual and auditory informativeness ($F_{3.2,61.1} = 58.3$; $P < 0.001$; Greenhouse–Geisser corrected). Similarly, for reaction times (limited to correct trials only), the 2-way RM ANOVA revealed significant main effects of vision ($F_{1.4,26.6} = 121.3$; $P < 0.001$) and audition ($F_{1.2,22.2} = 52.5$; $P < 0.001$) as well as an interaction between the 2 modalities ($F_{3.2,61.1} = 58.3$; $P < 0.001$) after Greenhouse–Geisser correction. Similar results were obtained when the (2-way RM) ANOVA excluded the 1) intact or 2) noise conditions (see Supplementary Material). These results show that categorization performance accuracy and reaction times depend on the informativeness of the auditory and visual modalities in an interactive fashion. In other words, the contribution of the auditory modality to object categorization depends on the informativeness of the visual modality and vice versa (see Supplementary Table 1). For instance, the effect of increasing auditory informativeness on object categorization is more pronounced when the visual stimulus is uninformative.

With respect to the functional imaging data, we focused specifically on 1) the bisensory and unisensory intact conditions (i.e., A_iV_i, A_nV_i, A_iV_n) and 2) the bisensory and unisensory degraded conditions (i.e., A_dV_d, A_nV_d, A_dV_n). The conditions A_nV_i, A_iV_n, A_nV_d, A_dV_n are referred to as unisensory to indicate that object information is provided only in one modality (see Fig. 1). In order to investigate crossmodal

benefits, we compared (paired t-tests) accuracy rates and reaction times of the bisensory condition with those of the best (i.e., most accurate and fastest) unisensory condition. A significant increase in performance accuracy was found only for the degraded stimuli (intact: $t_{19} = 0.5$; $P > 0.05$; degraded: $t_{19} = 3.4$; $P < 0.01$). A significant decrease in reaction times was found for both bisensory conditions (intact: $t_{19} = -3.3$; $P < 0.01$; degraded: $t_{19} = -4.6$; $P < 0.001$), with the benefit being greater for degraded stimuli (Fig. 2).

To further characterize subject's performance, signal sensitivity measures d' were computed for the degraded conditions only (see, Materials and Methods). Across subjects (mean ± standard error of the mean [SEM]), the d' of the auditory modality (A_dV_n d': 1.60 ± 0.15) was significantly greater than the d' of the visual modality (A_nV_d d': 0.90 ± 0.16; $t_{19} = -3.9$; $P < 0.01$). Furthermore, the d' of the bisensory condition (A_dV_d d': 2.16 ± 0.14) was significantly larger than the best unisensory d', showing a significant increase in perceptual sensitivity when both input modalities were present ($t_{19} = 3.1$; $P < 0.01$). Importantly, the d' of the bisensory condition (A_dV_d d': 2.16 ± 0.14) was significantly larger than d' predicted by the probability summation model (PSM_d d': 1.54 ± 0.16; $t_{19} = 3.7$; $P < 0.01$) with 16 of 20 subjects showing the effect. The prediction of the PSM is derived from the 2 unisensory d' sensitivity measures under the assumption that on each trial visual and auditory information are processed independently and combined in the behavioral decision using an either–or rule (Wickens 2002; see, Materials and Methods).

In summary, significant multisensory reaction time benefits were observed for both intact and degraded stimulus conditions. In contrast, a significant increase in performance accuracy was observed only for the degraded stimuli. Importantly, for degraded stimuli the increase in d' for the bisensory
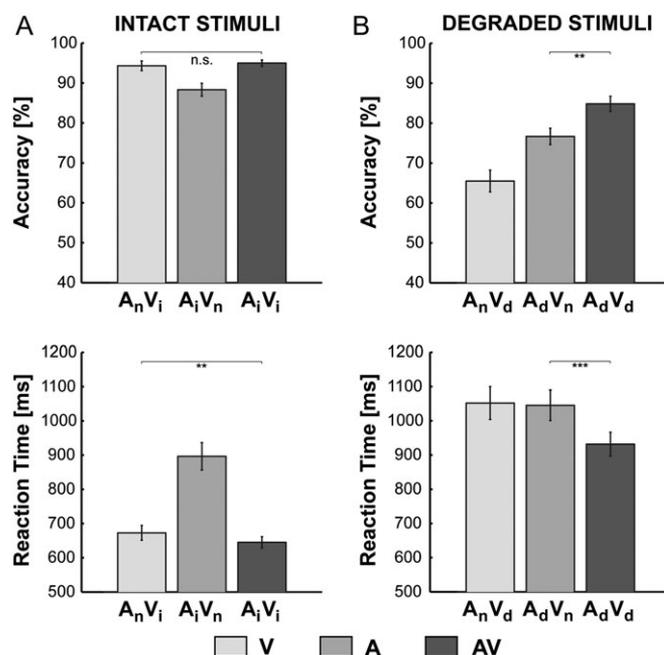

Figure 2. Categorization performance measures across subjects presented as bar plots for (A) intact, that is, unisensory (**V**: A_nV_i and **A**: A_iV_n) and bisensory (**AV**: A_iV_i) and (B) degraded, that is, unisensory (**V**: A_nV_d and **A**: A_dV_n) and bisensory (**AV**: A_dV_d) stimuli. Top: accuracy rates (% correct; mean ± SEM). Bottom: reaction times (mean ± SEM).

relative to the 2 unisensory conditions exceeded the prediction of the PSM. This suggests that subjects efficiently integrated visual and auditory information. On the other hand when the stimuli were intact, subjects may have relied primarily on the more informative, that is, the visual modality providing sufficient information for bisensory object categorization. However, the interpretation of the multisensory benefit for the intact conditions is limited by the fact that subjects achieved near-ceiling performance.

Target Detection Task

During the target detection 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 near-ceiling performance (mean ± SEM) for the target items with accuracies (visual: 93.3 ± 1.8%; auditory: 94.9 ± 1.8%; audiovisual: 95.4 ± 1.7%) and reaction times (visual: 461 ± 24 ms; auditory: 457 ± 27 ms; and audiovisual: 409 ± 23 ms) suggesting that they maintained bimodal attention. A 1-way RM ANOVA of the target item conditions (auditory, visual, and audiovisual) identified a significant main effect in terms of reaction times ($F_{1.7,32.9} = 9.6$; $P < 0.01$, Greenhouse–Geisser corrected) but not in terms of accuracy ($F_{1.4,26.7} = 0.8$; $P > 0.05$, Greenhouse–Geisser corrected). Post hoc comparisons (Bonferroni corrected) revealed faster audiovisual reaction times compared with auditory and visual reaction times, respectively, but no difference between the latter 2. During the target detection task, the total number of false alarms (i.e., responses to object stimuli) across subjects (mean ± SEM) was negligible (0.65 ± 0.25).

Neuroimaging Data

The data were analyzed twice: The first analysis assumed one underlying process that generates the profiles of BOLD and behavioral responses. The second analysis investigated whether there are differences in BOLD–responses that cannot be predicted by reaction times and hence included reaction times as an additional covariate of no interest to account for differences in processing times across conditions. As the 2 analyses provided nearly equivalent results, we report the results from the initial first analysis.

In both cases, the analysis was performed in 3 steps. First, we tested for superadditive and subadditive audiovisual interactions separately for intact and degraded stimuli. Second, we investigated the inverse effectiveness principle by directly comparing interactions for intact and degraded stimuli. Third, we investigated the functional relevance of superadditive and subadditive integration profiles by relating subjects' multisensory behavioral benefits to their audiovisual BOLD interactions in the degraded conditions. To increase contrast efficiency, all interactions were tested averaged across the 2 task conditions. However, to characterize the effects more thoroughly, we

Table 1
Effects of superadditivity, subadditivity, inverse effectiveness, and perceptual benefit

Regions	MNI coordinates			z score (peak)	P value$_c$ (cluster)	voxels	Categorization		Target detection	
	x	y	z				z score	P value (uncorr.)	z score	P value (uncorr.)
Superadditive interactions for degraded stimuli: $(A_dV_d - A_nV_n) - ([A_nV_d - A_nV_n] + [A_dV_n - A_nV_n])$										
R. superior frontal gyrus	24	21	66	4.02	0.004	56	4.11	0.000	2.61	0.005
R. superior medial frontal gyrus	9	36	63	3.99			3.90	0.000	2.65	0.004
L. angular gyrus	−42	−57	30	3.61	0.044	34	2.70	0.004	2.59	0.005
Subadditive interactions for degraded stimuli: $([A_nV_d - A_nV_n] + [A_dV_n - A_nV_n]) - (A_dV_d - A_nV_n)$										
L. insula (anterior)	−33	27	6	4.96	0.000	90	3.10	0.001	3.18	0.001
Subadditive interactions for intact stimuli: $([A_nV_i - A_nV_n] + [A_iV_n - A_nV_n]) - (A_iV_i - A_nV_n)$										
L. STS/STG (anterior)	−57	−9	−9	4.86	0.001	75	2.78	0.003	3.71	0.000
	−54	−12	−9	4.85			2.90	0.002	3.34	0.000
L. STS/STG (posterior)	−57	−45	9	3.77	0.013	44	1.95	0.026	2.62	0.004
	−57	−36	6	3.70			1.80	0.036	2.98	0.001
R. STS (posterior)	69	−42	9	4.48	0.000	228	3.20	0.001	2.33	0.010
R. STS (middle)	54	−21	−6	4.46			1.79	0.037	3.13	0.010
R. STG (anterior)	48	−3	−12	3.61			2.23	0.013	2.22	0.013
Inverse effectiveness principle: superadditivity$_{degraded}$ > superadditivity$_{intact}$										
L. STS (anterior)	−60	−9	−9	4.44	0.013	40	1.87	0.032	3.17	0.001
	−57	−6	−9	3.86			1.91	0.028	3.55	0.000
L. STS (posterior)	−60	−36	0	3.72	0.029#	17	1.60	0.056	3.05	0.001
	−60	−39	3	3.45			1.98	0.024	3.34	0.000
L. STS (middle)	−63	−33	0	3.37			1.76	0.040	2.65	0.004
R. STS/STG (posterior)	66	−33	9	4.22	0.004	50	3.07	0.001	3.52	0.000
	69	−39	9	4.21			3.05	0.001	3.63	0.000
Regression analysis for degraded stimuli: multisensory perceptual benefit versus AV–BOLD-response interaction										
L. precentral gyrus	−39	−21	54	4.24	0.001	67	2.26	0.012	2.02	0.022
L. postcentral gyrus	−42	−36	57	3.82			1.47	0.074	2.69	0.004
L. STS/STG (posterior)	−60	−42	12	3.84	0.041#	16	2.24	0.013	2.80	0.003
R. STS (posterior)	54	−33	6	3.64	0.006#	30	0.55	0.344	3.45	0.000
	60	−36	3	3.60			0.22	0.471	3.18	0.001
	48	−39	3	3.54			0.74	0.261	3.69	0.000

Note: P value$_c$, P value corrected at cluster level for multiple comparisons (whole-brain analysis, i.e., 54 518 voxels) or # the anatomical STS (MTG/STG/HG) search volume (3764 voxels). P value (uncorr.), uncorrected P value at peak voxels of clusters for the effects of interest within each task condition. L, left; R, right.

additionally report the effects for each task separately at the peak coordinates reported in Table 1 (note: the statistical inference is based only on the contrasts that pooled over tasks; the task-selective effects are reported only for descriptive purposes). Each effect was tested for in the whole brain and subsequently in our a priori volume of interest (i.e., encompassing HG, STS, and MTG bilaterally).

Superadditive Interactions

No superadditive interactions were observed for intact stimulus configurations corrected for multiple comparisons within the entire brain or the STS search mask. Superadditive (i.e., positive) interactions were found for degraded stimuli in the left angular gyrus and right superior and superior medial frontal gyrus. Within these areas, the summed activity of bisensory (i.e., A_dV_d) and low-level control conditions (i.e., A_nV_n) significantly exceeded the sum of the 2 unisensory conditions (i.e., $A_nV_d + A_dV_n$). Yet, the interactions in these regions are not discussed further, as in these areas 1) stimulus processing was associated with deactivations relative to fixation and 2) the low-level control condition A_nV_n induced the greatest activation (i.e., least deactivation) and was thus driving the superadditive interaction.

Subadditive Interactions

For intact stimulus configurations, prominent subadditive interactions were found along STS bilaterally. These effects were located in 1) the anterior and posterior portions of STS of the left hemisphere and 2) STG and middle/posterior STS of the right hemisphere (Fig. 3A). Within these areas, the sum of the unisensory BOLD–responses (i.e., A_nV_i and A_iV_n) significantly exceeded the sum of the bisensory (i.e., A_iV_i) and low-level control (i.e., A_nV_n) conditions. More precisely, the bisensory responses (i.e., A_iV_i) were consistently smaller than the responses elicited by the most effective unisensory stimulus (i.e., A_iV_n) indicating not only a subadditive but even suppressive form of multisensory interaction. These suppressive effects were particularly pronounced in the middle and anterior portions of the STS. As shown in the parameter estimate plots (Fig. 3B), the STS regions are auditory dominant, that is, unisensory auditory object stimuli elicited greater BOLD–responses than the unisensory visual object stimuli. However, in the context of an informative visual stimulus (i.e., bisensory stimulation), the auditory response is suppressed relative to its unisensory effect. Thus, the effect of the visual stimulus in STS is primarily of a modulatory nature in that it attenuates STS responses to auditory stimuli. Furthermore, at a lower statistical threshold, these suppressive effects could be shown both for the categorization and the target detection tasks suggesting that they may emerge at the perceptual rather than response selection level (Table 1; note: these task-selective effects were estimated with a lower degree of efficiency as they were based on fewer trials). It is important to note that not only subadditive but also suppressive interactions were observed for intact stimuli. While subadditive interactions could be due to ceiling effects in the BOLD–response, suppressive interactions are unlikely to emerge as a result of BOLD–response saturation alone.

For degraded stimulus configurations, subadditive interactions were found in the left anterior insula (Table 1) and the right superior colliculus ($x = 6$, $y = -33$, $z = -6$). However, the subadditive interaction effect in the right superior colliculus was only statistically significant ($z_{peak} = 3.59$; $P_{cluster} < 0.05$ with $k_{cluster} = 5$; $k_{ROI} = 192$) when correcting for multiple comparisons within a region of interest (ROI) that comprised the superior colliculi as similarly defined by other authors (Fairhall and Macaluso 2009).

The Effect of Stimulus Informativeness on the Multisensory Integration Profile

Brain regions obeying the principle of inverse effectiveness were identified by comparing superadditive BOLD–responses for degraded relative to intact stimuli. This statistical comparison revealed effects in anterior and posterior regions of the STS within the left hemisphere and STG and posterior STS of the right hemisphere (Fig. 3C). The clusters were partially overlapping with regions showing subadditive interactions for intact stimuli (Fig. 5 and Supplementary Fig. S3). At a lower threshold of significance, the peak voxels within these areas showed a pattern of inverse effectiveness consistently for both tasks (Table 1); yet, the effect was more reliably detected in the target detection task (i.e., passively attending the object stimuli). As shown in the parameter estimate plots (Fig. 3D), the regions exhibit strong subadditive interactions for intact stimulus configurations but additive to superadditive integration profiles for degraded stimuli. For example, the peak voxel in the left posterior STS (x, y, z coordinates: -60, -39, 3) showed subadditive interactions ($z = 2.92$; $P < 0.001$, uncorrected) for intact stimuli and a tendency toward a superadditive interaction for degraded stimuli ($z = 1.52$; $P = 0.067$, uncorrected). The remaining areas (e.g., left anterior STS; -60, -9, -9; Fig. 3D, upper panel) showed subadditive interactions for intact stimuli ($z = 3.93$; $P < 0.001$, uncorrected) but performed operations that were—even at an uncorrected threshold—not statistically different from an additive combination of the modality-specific inputs for the degraded stimulus configurations ($z = 0.71$; $P = 0.24$, uncorrected). Thus, degraded stimuli are primarily associated with additive response combinations in our study. As both intact and degraded bisensory stimuli were associated with shorter reaction times relative to their unisensory components, this activation pattern cannot easily be attributed to differences in processing time (i.e., reaction time) or attentional demands. Furthermore, it is unlikely to emerge from differences in accuracy, as errors and missed responses were modeled separately and not included in this comparison (though note that correct trials may also include guessed responses that are associated with increased uncertainty and are increased for degraded conditions). Furthermore, the effect was even more pronounced for the target detection task that did not require responses to the audiovisual object stimuli. Collectively, these results suggest that the profiles of higher order audiovisual integration in STS (i.e., additive responses vs. subadditive interactions) are dictated by the informativeness of the auditory and visual modalities in a similar way as they depend on physical stimulus intensity in the superior colliculus (Stanford et al. 2005).

The Effect of the Subject-Specific Multisensory Benefit on the Multisensory Integration Profile

The functional relevance of superadditive and subadditive integration profiles was investigated in a second-level regression analysis (within SPM) that used the subjects' multisensory behavioral benefit (i.e., d'(bisensory) – best(d'[unisensory])) as a predictor for their superadditive BOLD–responses (i.e., $(A_dV_d - A_nV_n) - ([A_nV_d - A_nV_n] + [A_dV_n - A_nV_n])$
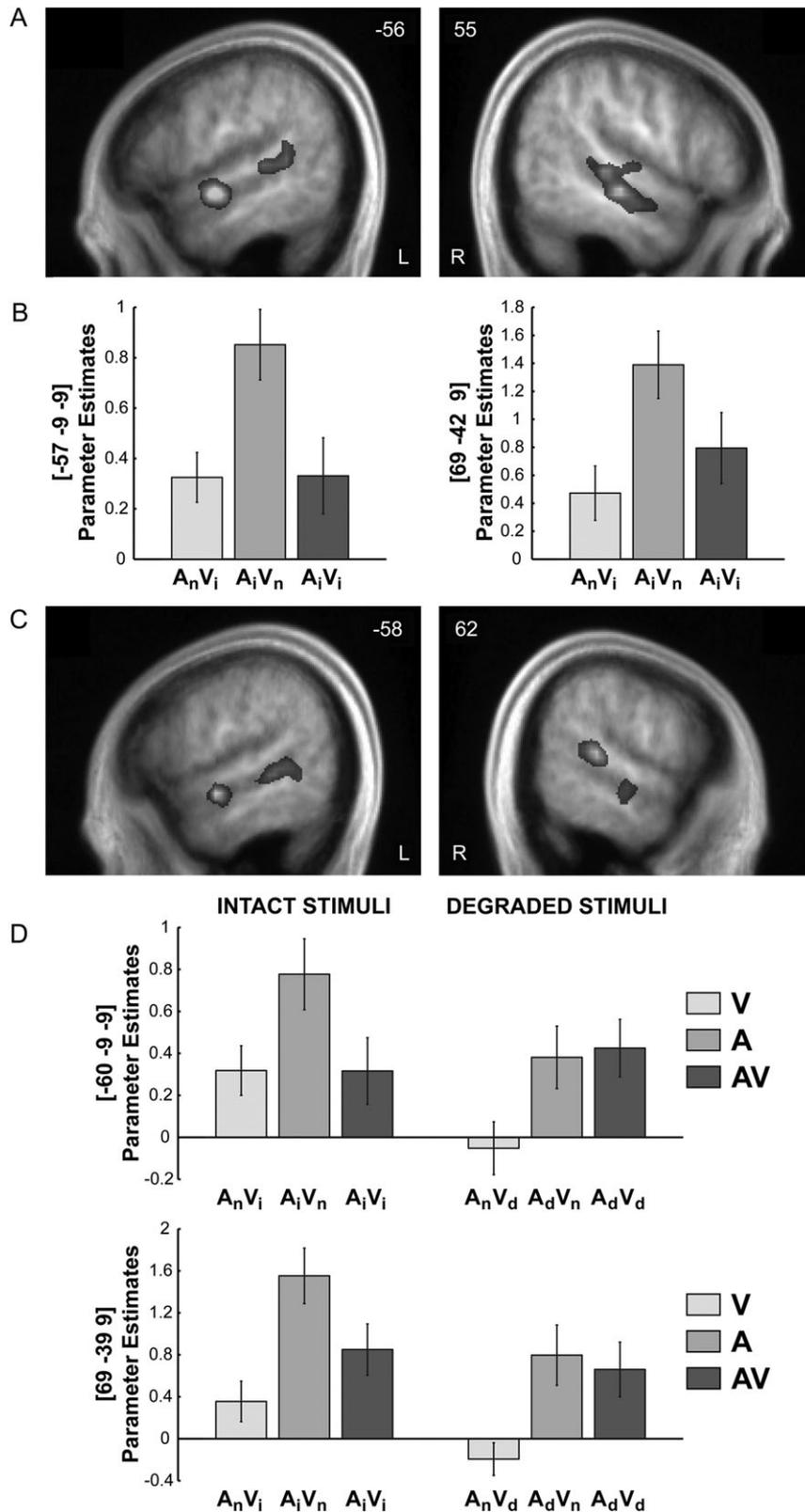
Figure 3. (*A*) Subadditive interactions for intact audiovisual stimulus configurations in left and right STS and STG on sagittal slices of a mean structural image created by averaging the subjects' normalized structural images. Height threshold: $P < 0.005$, uncorrected for illustrational purposes only. Extent threshold: >70 voxels. (*B*) Parameter estimates (mean $\pm$ SEM) for intact unisensory (**V**: A_nV_i and **A**: A_iV_n) and intact bisensory (**AV**: A_iV_i) conditions after subtraction of the low-level control condition (**N**: A_nV_n) at given coordinate locations. The bar graphs represent the size of the effect in nondimensional units (corresponding to % whole-brain mean). These effects are activations pooled (i.e., averaged) over task conditions. (*C*) Activations pertaining to inverse effectiveness in left and right STS and STG on sagittal slices of a mean structural image. Height threshold: $P < 0.005$, uncorrected for illustration purposes only. Extent threshold: >70 voxels. (*D*) Parameter estimates (mean $\pm$ SEM) for intact unisensory (**V**: A_nV_i and **A**: A_iV_n), intact bisensory (**AV**: A_iV_i), degraded unisensory (**V**: A_nV_d and **A**: A_dV_n), and degraded bisensory (**AV**: A_dV_d) conditions after subtraction of the low-level control condition (**N**: A_nV_n) at given coordinate locations. The bar graphs represent the size of the effect in nondimensional units (corresponding to % whole-brain mean). These effects are activations pooled (i.e., averaged) over task conditions.

interaction contrast). The superadditive BOLD–responses were averaged across tasks and limited to degraded stimulus conditions and correct responses only. The multisensory benefit positively predicted superadditive BOLD–responses in the right posterior STS, left posterior STG/STS, and left pre- and postcentral gyri (Fig. 4A). Figure 4B shows the effect of the audiovisual perceptual benefit on the superadditive BOLD–responses in left posterior STS: While subjects with no multisensory benefit exhibited subadditive and even suppressive audiovisual interactions, those with high multisensory benefits showed superadditive interactions (Fig. 4B). Even though, at a lower threshold of significance, multisensory interactions were predicted by the audiovisual benefit for both tasks at most of the reported peak voxels (Table 1), this effect was far more reliably found in the target detection task (in which subjects passively attended the object stimuli). In fact, an additional regression analysis limited to the target detection task revealed effects at the cluster level only in STS in the absence of effects in the pre- and postcentral gyri (see bottom Supplementary Table S2 for the results from the analysis that accounted for reaction time confounds). In contrast, no significant effects at the cluster level were observed, when limiting the analysis to the categorization task only. This additional analysis suggests a functional dissociation for STS and pre-/postcentral gyri in mediating multisensory benefits. We suggest that multisensory facilitation may emerge at multiple processing stages. In particular, multisensory integration may facilitate the emergence of a multisensory percept and response selection. During the target detection task, the audiovisual perceptual benefit (unconfounded by response selection processes) predicts superadditive BOLD–responses selectively in the posterior STS bilaterally, which may thus primarily reflect the emergence of a multisensory percept. In contrast, in pre-/postcentral gyri multisensory benefits predicted audiovisual interactions only when pooling over both tasks. This response profile may link the audiovisual interactions more closely with multisensory facilitation of response selection. The less significant effects in STS for the categorization task may have arisen due to differences in the analysis between the 2 tasks. In the categorization, task errors were modeled separately, while in the target detection task trials could not be sorted into correct and wrong responses, since subjects did not respond to the audiovisual object stimuli but only to the few target stimuli. Consequently, fewer trials entered into the estimation of the categorization response rendering the estimation of this contrast less efficient. In addition, the separation of correct and wrong responses may have partly modeled the behavioral gain during the categorization task. Regardless of the differences between the 2 tasks, it is important to emphasize that the predictive relationship between behavioral multisensory benefit and neural superadditivity is present in the target detection where subjects did not respond to the audiovisual object trials that entered into the analysis. This activation profile renders it unlikely that superadditive STS responses are purely attributable to differential demands in attentional resources or task-induced (e.g., response selection) processing. Instead, it suggests that superadditive responses mediate multisensory perceptual benefit during object processing.

To illustrate the relation between subjects' performance and neural activations, the perceptual sensitivity measures (Fig. 4D) and parameter estimates in left STS (Fig. 4E) are shown for A_nV_d, A_dV_n, A_dV_d separately for 3 groups of subjects that were categorized according to 1) no (benefit $\leq$ 0), 2) low (0 < benefit $\leq$ 1), or 3) high (benefit > 1) multisensory benefits. This classification of subjects highlights 3 main aspects: First, the multisensory benefit at the behavioral level (Fig. 4D) is clearly reflected in the multisensory integration profile in STS (Fig. 4E): Subadditive profiles (t_6 = 3.1; P < 0.05; one sample t-test) were associated with no benefit, additive (t_8 = −2.2; P > 0.05) profiles with a low benefit, and superadditive (t_3 = −2.7; P < 0.05; 1-tailed) profiles with a high multisensory perceptual benefit. Second, the increasing benefit at the behavioral level and the superadditivity at the neural level are driven by 2 main factors: 1) decreasing unisensory and 2) increasing bisensory d' measures and BOLD–responses. In fact, the close relationship between BOLD–response and behavior becomes even more apparent in an additional post hoc regression analysis (Fig. 4C) that uses the sum of the unisensory BOLD–responses in left STS to predict the perceptual multisensory benefit. Even though these 2 measures are independent (under the null hypothesis), the perceptual benefit is indeed significantly predicted by the sum of the 2 unisensory BOLD–responses in STS (b_{sl} = −0.24; R^2 = 0.22; $P_{uncorrected}$ < 0.05; note: this additional analysis was performed only for explanatory purposes and is not statistically independent from our main initial analysis). Third, in line with the maximum likelihood model of multisensory integration, the behavioral benefit is maximal when the reliabilities of the unisensory inputs (as indicated by the d') are equal (i.e., Fig. 4D, right subplot; Ernst and Banks 2002; Alais and Burr 2004; Ross et al. 2007; Morgan et al. 2008). Thus, the perceptual benefit observed in our study is maximal when 1) the unisensory reliabilities (as indexed by d') are low and 2) the ratio of subject-specific unisensory reliabilities is close to 1. Future studies will need to clearly dissociate the roles of unisensory reliabilities and their ratio on behavioral benefit and superadditive BOLD–responses.

Discussion

This study was designed to identify and characterize the neural systems and mechanisms mediating the integration of higher order audiovisual object features in recognition and categorization. In brief, our experiment demonstrates higher level audiovisual integration effects selectively in the STS bilaterally. Consistent with the inverse effectiveness principle, these multisensory interactions are primarily subadditive for intact stimuli but turn into additive (with a trend to superadditive) effects for degraded, near-threshold stimuli. Importantly, when holding stimulus degradation constant, superadditive responses in STS are predicted by subjects' perceptual multisensory benefits suggesting a functional role for the distinct profiles of multisensory integration at the neuronal level.

While it is well established that auditory and visual information converges in subcortical (Wallace et al. 1996; Calvert et al. 2001), putatively unisensory (Macaluso et al. 2000; Murray et al. 2005; Lehmann et al. 2006; Watkins et al. 2006; Meienbrock et al. 2007), and higher level association regions (Calvert et al. 2000; Wright et al. 2003; Amedi et al. 2005; Saito et al. 2005) during speech and object recognition, the types of information (e.g., spatiotemporal or object related) that are integrated in this multitude of integration sites remain unclear (Driver and Noesselt 2008). Controlling for integration
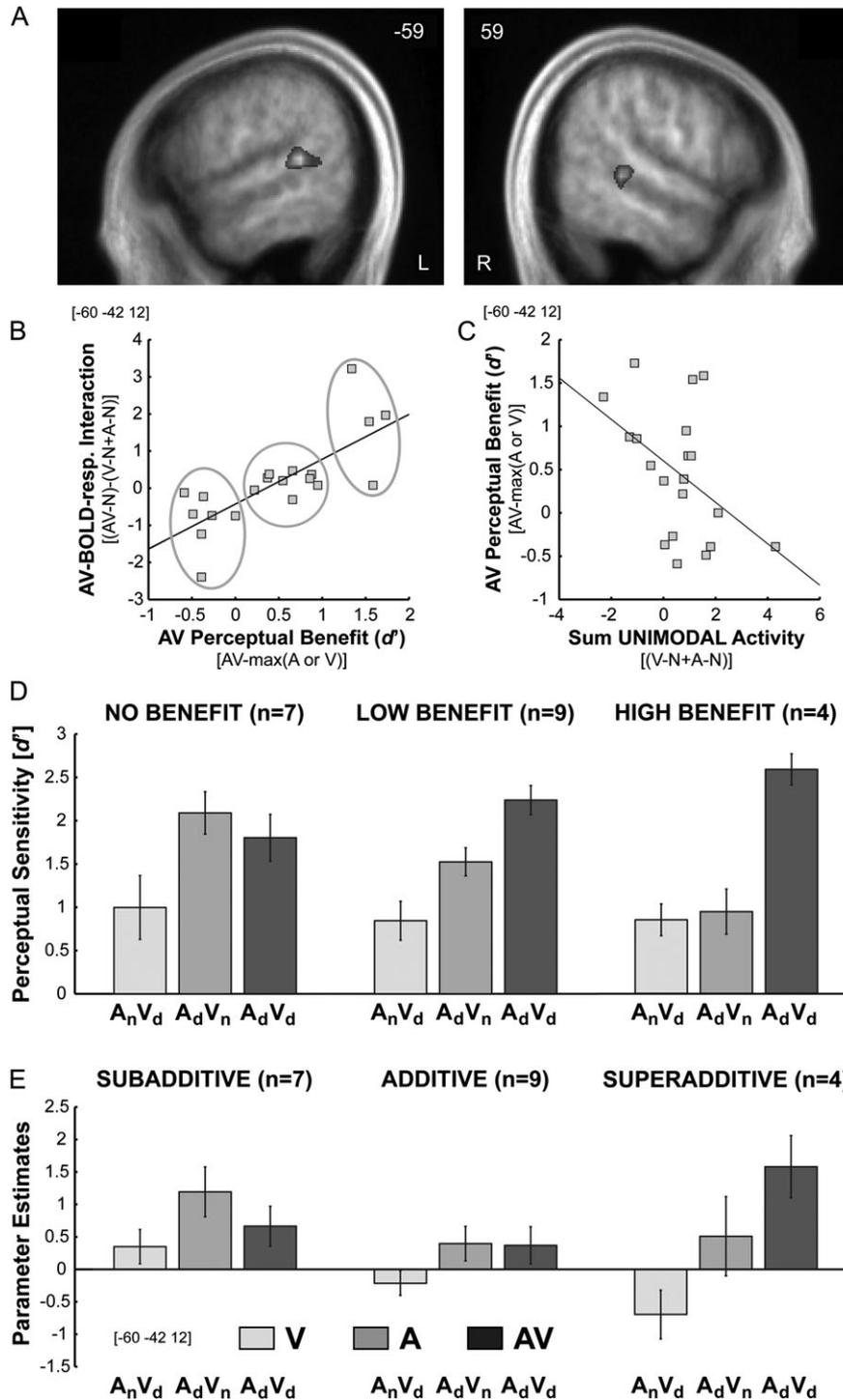
Figure 4. (A) Effects of subjects' multisensory perceptual benefit on superadditive BOLD–response interactions in left and right posterior STS for degraded stimuli on sagittal slices of a mean structural image. Height threshold: $P < 0.005$, uncorrected for illustrational purposes. Extent threshold: >70 voxels. (B) Scatter plot depicting the regression of the superadditive interaction contrast (ordinate) on the perceptual benefit (abscissa) across subjects at ($x = -60$, $y = -42$, $z = 12$). Superadditivity and perceptual benefit were computed only for the degraded conditions: bisensory (**AV**: A_dV_d), unisensory (**V**: A_nV_d and **A**: A_dV_n), and the low-level control (**N**: A_nV_n) conditions (for the fMRI data averaged across task contexts). The across subjects means (standard deviation [SD]) were 0.13 (1.19) for the superadditive interaction effects and 0.46 (0.74) for the perceptual benefits. The circle indicates the 3 groups of subjects with no (benefit ≤ 0), low (0 < benefit ≤ 1), and high (benefit > 1) behavioral benefit. (C) Scatter plot depicting the regression of the multisensory perceptual benefit (ordinate) on the (summed) unisensory BOLD–responses (abscissa) across subjects at ($x = -60$, $y = -42$, $z = 12$). Perceptual benefit and (summed) unisensory BOLD–responses were computed only for the degraded bisensory (**AV**: A_dV_d), unisensory (**V**: A_nV_d and **A**: A_dV_n), and the low-level control (**N**: A_nV_n) conditions. The (summed) unisensory BOLD–responses (relative to the low-level control condition) are averaged across task contexts. The across subjects means (SD) were 0.46 (0.74) for the perceptual benefits and 0.59 (1.44) for (summed) unisensory BOLD–responses. (D) Perceptual sensitivity d' (mean ± SEM) for degraded bisensory (**AV**: A_dV_d) and unisensory (**V**: A_nV_d and **A**: A_dV_n) conditions for the 3 groups of subjects with no (benefit ≤ 0), low (0 < benefit ≤1), or high (benefit > 1) behavioral benefit. (E) Parameter estimates (mean ± SEM) at ($x = -60$, $y = -42$, $z = 12$) for degraded bisensory (**AV**: A_dV_d) and unisensory (**V**: A_nV_d and **A**: A_dV_n) conditions after subtraction of the low-level control (**N**: A_nV_n) for groups of subjects with no (benefit ≤ 0), low (0 < benefit ≤ 1), and high (benefit > 1) behavioral benefit. The bar graphs represent effect sizes in nondimensional units (corresponding to % whole-brain mean) averaged across task contexts.
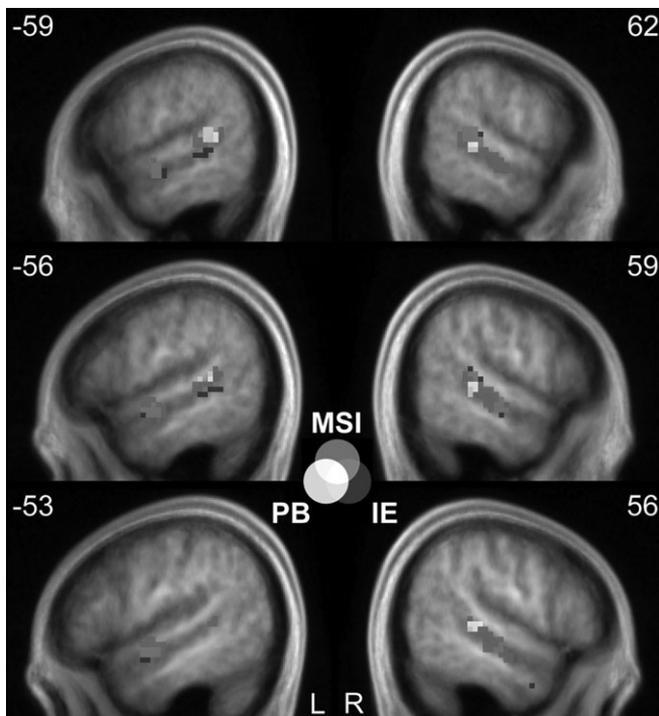
Figure 5. Overview figure illustrating the overlap of effects of multisensory integration, inverse effectiveness, and perceptual benefit (PB) on sagittal slices of a mean structural image created by averaging the subjects' normalized structural images. $P < 0.05$, corrected for spatial extent (auxiliary height threshold: $P < 0.001$, uncorrected; as displayed in Table 1). Gray 55%: subadditive multisensory interactions (MSIs) for intact stimuli. Gray 80%: activations pertaining to inverse effectiveness (IE) principle. Gray 20%: effects of subjects' multisensory PB on superadditive BOLD–response interactions. Gray 50%: (MSI) and (IE). Gray 5%: (MSI) and (PB). White: (MSI) and (IE) and (PB). See Supplementary Figure S3 for a colored version of this figure.

processes of spatiotemporal information and low-level stimulus features, our experimental paradigm revealed higher order audiovisual interactions selectively in posterior, middle, and anterior portions of the STS, bilaterally. These results suggest that distinct types of stimulus parameters may be integrated selectively at different levels of the sensory hierarchy with higher order association cortex (i.e., STS) sustaining integration of audiovisual information about an object's category or identity.

In terms of neural mechanisms, neurophysiological studies have previously shown the coexistence of subadditive, additive, and superadditive neuronal populations (Laurienti et al. 2005; Perrault et al. 2005), with the profile of multisensory integration being determined by stimulus efficacy that is usually manipulated by low-level stimulus features (Stanford et al. 2005). Indeed, numerous neurophysiological studies have demonstrated that low-level stimulus features such as physical intensity or contrast determine the integration profiles within the superior colliculus (Stein and Stanford 2008). However, it is still unclear whether higher order features such as stimulus informativeness can play a similar role. Our results show that higher order stimulus properties such as degradation level or informativeness dictate the multisensory integration profiles in STS: Subadditive audiovisual interactions in the STS for intact stimuli turn into additive (with a trend to superadditive) integration profiles for degraded, near-threshold stimuli (Stevenson and James 2009). Importantly, the audiovisual interactions were not only subadditive but also suppressive, that is,

the response to bisensory stimuli was less than the response to the most effective unisensory stimulus, which cannot easily be explained by saturation or other nonlinearities in the BOLD–response (Buxton et al. 2004). Suppressive interactions have previously been attributed to spatial, temporal, or semantic incongruencies (Wallace et al. 1996; Calvert et al. 2000, 2001). As our stimuli combined a static image with an inherently dynamic sound, they may have elicited perceptual incongruency to a certain degree. Thus, future studies need to generalize our findings to more environmentally valid audiovisual stimuli such as movies. Nevertheless, even though the stimuli were not dynamically synchronous, they were aligned with respect to audiovisual stimulus onsets and offsets. Furthermore, both intact and degraded conditions used pictures and sounds; yet, suppressive interactions were found selectively for the intact stimulus conditions rendering perceptual incongruency as the primary explanatory mechanism rather unlikely. Instead, our results converge with recent neurophysiological studies showing pronounced suppressive multisensory interactions in the absence of incongruency manipulations. For instance, in the anterior ectosylvian cortex of the cat, the response to somatosensory input was significantly modulated, more specifically suppressed, by additional auditory input (Dehner et al. 2004). Similarly, in the macaque STS, ventrolateral prefrontal and ventral intra-parietal cortex, both MSE and suppression have been reported (Barraclough et al. 2005; Sugihara et al. 2006; Avillac et al. 2007). Even though suppressive effects at the neuronal level may not necessarily translate into suppressive BOLD–response interactions (Kayser et al. 2009), this series of studies demonstrates the existence of suppressive interactions in the absence of incongruency manipulations.

Even though the inverse effectiveness principle has played a fundamental role in multisensory research over the past decade (Stanford and Stein 2007; Stein et al. 2009), direct evidence for its functional relevance and relationship to subjects' behavior has been elusive (Grant and Walden 1996; Holmes 2007, 2009; Ross et al. 2007). The present study provides compelling evidence for the behavioral relevance of the inverse effectiveness principle in object categorization at 2 levels: First, manipulating stimulus informativeness or degradation, our study maps different behavioral performances on additive and suppressive profiles of integration: Suppressive integration profiles are observed for intact conditions that are associated only with faster categorization of bisensory relative to unisensory stimuli. In contrast, additive integration profiles are found for degraded stimuli that are associated both with faster and more accurate categorization performance. Indeed, only for the degraded stimuli did accuracy exceed the performance predicted by the PSM (i.e., visual and auditory information are processed independently). These findings provide initial evidence that performance accuracy may be related to superadditivity. Second and more convincingly, when holding stimulus informativeness constant (i.e., unconfounded by stimulus differences) and focusing only on the degraded stimuli, regression analyses demonstrate that subjects' multi-sensory perceptual benefit (as measured by d') directly predicts their integration profiles selectively in STS. While subjects that do not benefit from audiovisual integration (i.e., high unisensory performance) show subadditive and even suppressive audiovisual interactions, subjects with great perceptual benefits (i.e., low unisensory performance) exhibit

superadditive integration profiles. Collectively, these results show that (only) in the posterior STS, the multisensory integration profiles and their inverse relationships (between unisensory responses and superadditivity) are functionally relevant for categorization performance: Superadditive interactions in posterior STS are associated with and may induce multisensory perceptual benefit and increased recognition performance.

Similar parallels between behavioral and neuronal measures have previously been reported for repetition priming: While priming-induced activation decreases are observed for intact stimuli that are more efficiently, that is, faster processed on repeated exposures, activation increases are shown for degraded stimuli that become recognizable through priming (Dolan et al. 1997; George et al. 1999; Doniger et al. 2001; Henson 2003). Similar to priming effects, the subadditive (or even suppressive) and superadditive profiles of multisensory integration may serve 2 aims: In the context of near-threshold unisensory inputs that are not recognized alone, bisensory stimulation decreases the thresholds of detection and identification enabling the emergence of higher order object representations and better categorization performance. In contrast, if a stimulus can easily be recognized at least in one modality as during near-ceiling performance for intact stimuli or for subjects with no perceptual benefit, suppressive interactions may reflect more efficient and faster processing through the dynamic weighting of the unisensory contributions according to their informativeness. These suppressive crossmodal effects may be mediated directly through the influence from predominantly visual areas. Alternatively, they may be mediated via top-down modulation from a frontoparietal attentional system.

In conclusion, using a novel elaborate interaction design, we were able to dissociate distinct multisensory response profiles underlying the integration of higher order audiovisual object features in the STS bilaterally. Consistent with the inverse effectiveness principle, the audiovisual integration profile was dictated by stimulus efficacy that depended in STS on the informativeness of the sensory-specific inputs: Suppressive interactions for bisensory intact stimuli turned into additive interactions for degraded, near-threshold stimuli. Crucially, the distinct integration profiles were functionally relevant and paralleled behavioral indices of MSE: Subjects with multisensory perceptual benefit showed superadditive interactions, while those that did not benefit from audiovisual stimulation exhibited suppressive interactions. The additive and superadditive multisensory interactions may mediate efficient integration of near-threshold inputs from multiple senses. In contrast, suppressive interactions may reflect more efficient processing when the stimulus can already be recognized in at least one modality.

Supplementary Material

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

Funding

Notes

Address correspondence to Sebastian Werner, Max Planck Institute for Biological Cybernetics, Spemannstrasse 41, 72076 Tübingen, Germany. Email: sebastian.werner@tuebingen.mpg.de.

References

Alais D, Burr D. 2004. The ventriloquist effect results from near-optimal bimodal integration. Curr Biol. 14:257–262.

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.

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

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

Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA. 2007. Neural basis of the ventriloquist illusion. Curr Biol. 17:1697–1703.

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

Buxton RB, Uludag K, Dubowitz DJ, Liu TT. 2004. Modeling the hemodynamic response to brain activation. Neuroimage. 23:S220–S233.

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, Lewis JW. 2004. Hemodynamic studies of audiovisual interactions. In: Calvert GA, Spence C, Stein BE, editors. The handbook of multisensory processes. Cambridge (MA): MIT Press. p. 483–502.

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.

Dehner LR, Keniston LP, Clemo HR, Meredith MA. 2004. Cross-modal circuitry between auditory and somatosensory areas of the cat anterior ectosylvian sulcal cortex: a 'new' inhibitory form of multisensory convergence. Cereb Cortex. 14:387–403.

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

Doniger GM, Foxe JJ, Schroeder CE, Murray MM, Higgins BA, Javitt DC. 2001. Visual perceptual learning in human object recognition areas: a repetition priming study using high-density electrical mapping. Neuroimage. 13:305–313.

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

Ernst MO, Banks MS. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. Nature. 415:429–433.

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.

Friston KJ, Holmes AP, Price CJ, Buchel C, Worsley KJ. 1999. Multisubject fMRI studies and conjunction analyses. Neuroimage. 10:385-396.

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, 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.

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, 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.

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.

Grant KW, Walden BE. 1996. Evaluating the articulation index for auditory-visual consonant recognition. J Acoust Soc Am. 100:2415-2424.

Henson RNA. 2003. Neuroimaging studies of priming. Prog Neurobiol. 70:53-81.

Holmes NP. 2007. The law of inverse effectiveness in neurons and behaviour: multisensory integration versus normal variability. Neuropsychologia. 45:3340-3345.

Holmes NP. 2009. The principle of inverse effectiveness in multisensory integration: some statistical considerations. Brain Topogr. 21:168-176.

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.

Kayser C, Petkov CI, Logothetis NK. 2009. Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. Hear Res. doi:10.1016/j.heares.2009.02.011.

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 ed.

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, 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, et al. 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.

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.

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.

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.

Macaluso E, Frith CD, Driver J. 2000. Modulation of human visual cortex by crossmodal spatial attention. Science. 289:1206-1208.

Meienbrock A, Naumer MJ, Doehrmann O, Singer W, Muckli L. 2007. Retinotopic effects during spatial audio-visual integration. Neuropsychologia. 45:531-539.

Meredith MA. 2002. On the neuronal basis for multisensory convergence: a brief overview. Brain Res Cogn Brain Res. 14:31-40.

Meredith MA, Stein BE. 1983. Interactions among converging sensory inputs in the superior colliculus. Science. 221:389-391.

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.

Morgan ML, DeAngelis GC, Angelaki DE. 2008. Multisensory integration in macaque visual cortex depends on cue reliability. Neuron. 59:662-673.

Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ. 2005. Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex. 15:963-974.

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

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

Perrault TJ, Vaughan JW, Stein BE, Wallace MT. 2003. Neuron-specific response characteristics predict the magnitude of multisensory integration. J Neurophysiol. 90:4022-4026.

Perrault TJ, Vaughan W, Stein BE, Wallace MT. 2005. Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. J Neurophysiol. 93:2575-2586.

Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. 2007. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environment. Cereb Cortex. 17:1147-1153.

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 audio-visual 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. Brain Res Cogn Brain Res. 14:187-198.

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

Stanford TR, Stein BE. 2007. Superadditivity in multisensory integration: putting the computation in context. Neuroreport. 18:787-792.

Stein BE, Meredith MA. 1993. The merging of the senses. Cambridge (MA): The MIT Press.

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

Stein BE, Stanford TR, Ramachandran R, Perrault TJ, Rowland BA. 2009. Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. Exp Brain Res. 198:113-126.

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.

Stevenson RA, Kim S, James TW. 2009. An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. Exp Brain Res. 198:183-194.

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.

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

Watkins S, Shams L, Tanaka S, Haynes JD, Rees G. 2006. Sound alters activity in human V1 in association with illusory visual perception. Neuroimage. 31:1247-1256.

Wickens TD. 2002. Multidimensional stimuli. Elementary signal detection theory. New York: Oxford University Press. p. 172-194.

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.