Audiovisual Synchrony Enhances BOLD Responses in a Brain Network Including Multisensory STS While Also Enhancing Target-Detection Performance for Both Modalities

Jennifer L. Marchant,[1,2]* Christian C. Ruff,[1,2,3] and Jon Driver[1,2]

[1]Wellcome Trust Centre for Neuroimaging at UCL, Institute of Neurology, University College London, London, WC1N 3BG, United Kingdom
[2]UCL Institute of Cognitive Neuroscience, University College London, London, WC1N 3AR, United Kingdom
[3]Laboratory for Social and Neural Systems Research, Institute for Empirical Research in Economics, University of Zurich, CH-8006 Zurich, Switzerland

◆ ═══════════════════════════════════ ◆

Abstract: The brain seeks to combine related inputs from different senses (e.g., hearing and vision), via multisensory integration. Temporal information can indicate whether stimuli in different senses are related or not. A recent human fMRI study (Noesselt et al. [2007]: J Neurosci 27:11431–11441) used auditory and visual trains of beeps and flashes with erratic timing, manipulating whether auditory and visual trains were synchronous or unrelated in temporal pattern. A region of superior temporal sulcus (STS) showed higher BOLD signal for the synchronous condition. But this could not be related to performance, and it remained unclear if the erratic, unpredictable nature of the stimulus trains was important. Here we compared synchronous audiovisual trains to asynchronous trains, while using a behavioral task requiring detection of higher-intensity target events in either modality. We further varied whether the stimulus trains had predictable temporal pattern or not. Synchrony (versus lag) between auditory and visual trains enhanced behavioral sensitivity (d') to intensity targets in either modality, regardless of predictable versus unpredictable patterning. The analogous contrast in fMRI revealed BOLD increases in several brain areas, including the left STS region reported by Noesselt et al. [2007: J Neurosci 27:11431–11441]. The synchrony effect on BOLD here correlated with the subject-by-subject impact on performance. Predictability of temporal pattern did not affect target detection performance or STS activity, but did lead to an interaction with audiovisual synchrony for BOLD in inferior parietal cortex. *Hum Brain Mapp 33:1212–1224, 2012.* © **2011 Wiley Periodicals, Inc.**

Key words: fMRI; multisensory; auditory; visual; superior temporal sulcus; asynchrony; predictability

◆ ═══════════════════════════════════ ◆

INTRODUCTION

Events in the environment often stimulate more that one sense. A burgeoning literature illustrates that the brain can exploit relations between stimuli in different senses to enhance sensory representations, via multisensory integration [e.g., for overviews see Beauchamp, 2005; Calvert et al., 2004; Doehrmann and Naumer, 2008; Driver and Noesselt, 2008; Ghazanfar and Schroeder, 2006; Kayser et al., 2009; Macaluso and Driver, 2005; Spence and Driver, 2004; Stein and Meredith, 1993]. Ideally multisensory integration should only combine information from different senses when this information is related. A variety of cues can indicate whether stimuli from different senses are related, such as their relative spatial [e.g., Macaluso et al., 2000, 2004; Stein and Meredith, 1993; Wallace et al., 1996], semantic [e.g., Beauchamp et al., 2004a,b; Ghazanfar et al., 2005; Hein et al., 2007; Noppeney et al., 2008], or temporal properties [e.g., Bischoff et al., 2007; Bushara et al., 2001; Calvert et al., 2001; Macaluso et al., 2004; Meredith et al., 1987; Noesselt et al., 2007; Stevenson et al., 2010; Wallace et al., 1996]. Here we focus specifically on temporal relations and primarily on human fMRI studies.

Human behavioral studies indicate that synchrony between auditory and visual events can potentially enhance their perceived saliency, compared to unisensory or asynchronous stimulation [e.g., Frassinetti et al., 2002; Lovelace et al., 2003; Odgaard et al., 2003, 2004; Stein et al., 1996]. These findings compliment an abundance of invasive animal studies demonstrating that the relative timing of events in different modalities can be a key determinant of whether and how multisensory integration arises between them in the brain [e.g., Kayser et al., 2008; Meredith et al., 1987; Stein and Wallace, 1996; Wallace et al., 1996]. Although there have been several multisensory human fMRI studies of spatial and/or semantic relations between stimuli in different senses [e.g., see Calvert et al., 2000; Doehrmann and Naumer, 2008; Macaluso and Driver, 2005, among many others], there have been somewhat fewer multisensory human fMRI investigations on the role of timing [though see Bushara et al., 2001; Calvert at al., 2001; Hertz and Amedi, 2010; Noesselt et al., 2007; Stevenson et al., 2010; van Atteveldt et al., 2007].

Noesselt et al. [2007] acquired human fMRI data during non-semantic trains of beeps and flashes with erratic jittered timing. Their key manipulation was whether the auditory and visual trains were synchronous or unrelated in timing (while conserving the same overall temporal statistics for each train in both conditions). They found that several brain regions were affected, but particularly highlighted a region (peak at $x = -54$, $y = -50$, $z = 8$) in multisensory left posterior STS [cf., Beauchamp et al., 2004a,b; Bischoff et al., 2007; Calvert et al., 2001; Hein et al., 2007; Macaluso et al., 2004; Meienbrock et al., 2007; Stevenson and James, 2009; Stevenson et al., 2010; Werner and Noppeney, 2010] that showed higher BOLD signal during the synchronous versus temporally unrelated audi-

tory-visual stimulus trains. They proposed that this region may serve to detect synchrony between auditory and visual stimuli [see also Macaluso et al., 2004; Stevenson et al., 2010].

Several new questions arise from these findings, which we addressed here. First there is the issue of how neural effects of audiovisual synchrony may relate to behavioral effects. Noesselt et al. [2007] were unable to relate their observed fMRI effects of audiovisual synchrony, on brain areas such as STS, to any behavioral impact of such synchrony. Here we introduced target monitoring tasks that had to be performed concurrently on the auditory and visual trains of stimuli, allowing us to determine if audiovisual synchrony affected target detection in either modality, and whether there was any relation of this to the fMRI effects.

A second issue is whether the temporal predictability of auditory and visual trains may matter for the impact of audiovisual synchrony on brain activity and on performance. There is a growing literature on so-called "predictive coding" in the study of perception [see Dayan et al., 1995; Friston, 2005; Helmholtz, 1860; Mumford, 1992; Rao, 1999] which emphasizes that predictable stimuli may be processed differently from unpredictable ones [e.g., den Ouden et al., 2009; Furl et al., 2010; Overath et al., 2007; Summerfield and Koechlin, 2008]. It has also been suggested that the brain may seek to derive the "generative model" [e.g., Friston, 2005] that can most readily explain and even predict sensory observations. In this respect, it may be noteworthy that in the Noesselt et al. [2007] audiovisual study, each train of stimuli was highly erratic in timing. As those authors noted, this made it highly unlikely that the two modalities would coincide accidentally, unless they were generated by common supramodal events in the external world. Thus the combination of unpredictable timing with perfect audiovisual synchrony may have provided particularly strong information that events in the two modalities were related.

But since only erratic, temporally unpredictable stimulus trains were used in Noesselt et al. [2007], they were unable to determine if this was actually critical for the observed influence on STS. One possibility is that activity for STS (and related areas) may be increased by audiovisual synchrony in a strictly bottom-up manner, regardless of predictable or unpredictable contexts. An alternative possibility is that audiovisual synchrony may have more impact for temporally unpredictable stimulus trains [as tested by Noesselt et al., 2007] than for predictable trains, since the latter have a simpler underlying "generative model." Note that arranging audiovisual synchrony for temporally regular trains only requires the first events in such trains to be aligned across the two modalities, since co-incident timing of all the subsequent events will then take care of itself due to the regularity; whereas for erratic trains every successive event needs to be specifically arranged in order to maintain synchrony across modalities.

Here we addressed these issues by manipulating not only audiovisual synchrony versus asynchrony for stimulus trains, but also the temporal predictability of those trains.

We examined the impact of these factors (and also any interaction between them) not only for fMRI activations, but also for performance in a target-monitoring task performed for both vision and hearing. The task was to detect higher-intensity targets occurring in either modality, within trains of non-target events at a standard intensity.

MATERIALS AND METHODS

Participants

Sixteen right-handed participants (mean age 24.7 years, 9 female) with no history of neurological or psychiatric illness gave written informed consent. All had normal or corrected-to-normal vision and normal hearing by self-report. This study was approved by the University College London Research Ethics Committee and conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).
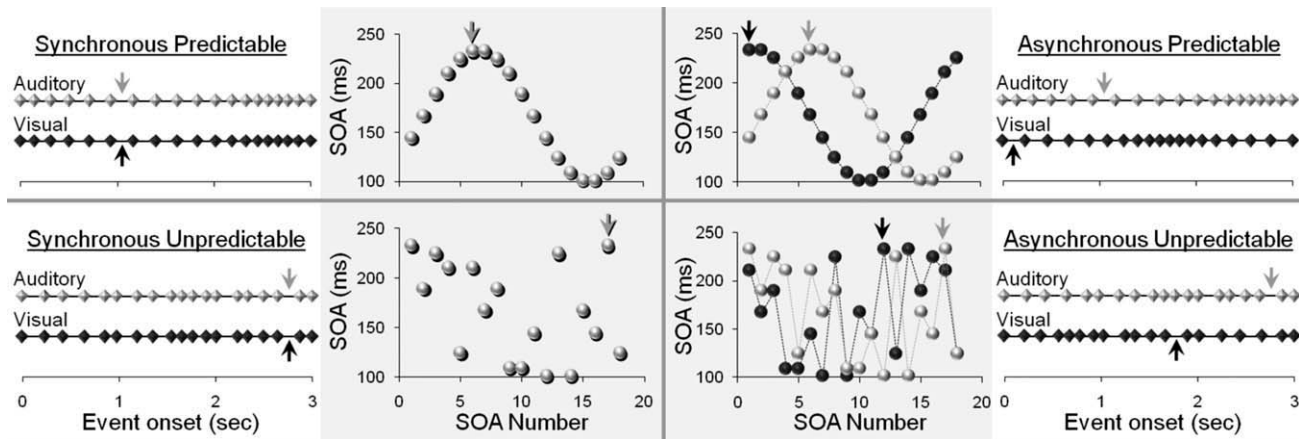
Stimuli

Each standard visual stimulus was a checkerboard (4.5° × 4.5°; 9 × 9 squares; light squares 6.40 cd mm^{-2}, dark squares 1.98 cd mm^{-2}) flashed for 33 ms in the upper right quadrant (centered at 8°) of a dark grey screen (1.37 cd mm^{-2}). A 30% of trials included a visual target that was identical except for having a higher contrast (exact value titrated for each participant as explained below, with brighter squares having a mean of 11.0 ± 0.24 cd mm^{-2}, darker squares, 0.22 ± 0.28 cd mm^{-2}; contrast change from standard averaging 6.34 ± 0.32 cd mm^{-2}). The standard auditory stimulus was a 33-ms 1,000-Hz pure tone [mean 49.5 ± 7.8 dB(A)] with 5-ms ramp applied at both the onset and offset. On 30% of trials (separate from those with a visual target) an auditory target was presented that was identical except louder. For auditory stimuli it was the standard stimulus intensity that was titrated for each participant. The auditory target was always presented at a maximal 60 dB(A); the difference in intensity from the titrated standard averaged 10.5 ± 7.8 dB(A). Actual stimulus intensities were set for each participant during in-situ practice with the scanner running, so that target hit-rate was ∼ 70% in each modality. Please note that the rationale underlying this task of monitoring for higher-intensity targets among standards was to provide a performance measure that might be influenced by our manipulations of audiovisual synchrony and/or temporal predictability for the trains. It was not our intention to identify "oddball-like" brain responses, and our higher-intensity targets were not particularly rare (occurring on 60% of trials when summed over modalities, see below). Instead we were interested in the possible impact of audiovisual synchrony and temporal predictability for the trains in which targets could appear, both for affecting target detection performance, and also for fMRI activations.

Each 3-s trial comprised presentation of a rapid auditory stimulus train containing 19 pure tones, plus a rapid visual stimulus train containing 19 checkerboard flashes. Both trains always started and ended at the same time for all conditions, to prevent any extended periods with only unimodal stimulation from arising during the asynchronous conditions, and to ensure that the initial-onset, final-offset, and duration was identical for all trains in all experimental conditions. (Please note that despite this close matching of conditions, we still found substantial behavioral and fMRI effects of audio-visual synchrony versus asynchrony throughout the train, see below). The same 18 stimulus-onset-asynchronies (SOAs) between successive events within each sensory train were used for each trial, so as to conserve temporal structure between modalities and conditions. These 18 SOAs were each multiples of the screen refresh rate 60 Hz, in the range 100–234 ms, derived from a sinusoid-like distribution around the mean SOA of 167 ms; see Figure 1. The SOAs on each trial were always fully sampled (once each) from this same underlying distribution of 18 SOAs, for both modalities.

Experimental Design

A 2 × 2 factorial design manipulated synchrony between the auditory and visual trains, and orthogonally the temporal "predictability" of events within those trains. Audiovisual (a) synchrony was manipulated via the relative start points in the SOA sequence for the two sensory trains (see Fig. 1). The start points were either the same for both trains (synchronous) or one train started 5–13 positions (550–1,500 ms) further along the selected sequence for one modality, then cycled around the earlier positions later (asynchronous) in a wrap-around design.

The predictable and unpredictable temporal-structure conditions were created by manipulating the order of the SOAs within a particular trial. The order of the SOAs could cycle around the sinusoidal structure already explained for successive SOAs (thus predictably), albeit starting at a random point in this cycle on each trial so as to match the distribution of possible starting SOAs from the unpredictable condition. Alternatively the order of the 18 SOAs was randomized (unpredictable). The same underlying SOA sequence was used for both the auditory and visual trains on any given trial, i.e., either both were predictable or both were unpredictable (with the specific unpredictable sequence on a given trial in the latter case being used either in a synchronous or asynchronous manner between the two modalities). Accidental coincidences between auditory and visual events in the "asynchronous" conditions (which actually presented the same patterns of successive SOAs, but with lags of 550–1,500 ms, see above) were not artificially prevented, but instead allowed to occur naturally at a rate (5.8% of events) which was too rare to analyze. Moreover, as will be seen, we found clear main effects of audiovisual synchrony for both behavioral

Figure 1.

Schematic illustration of how SOAs and train start-points were selected to yield the four conditions of our two (synchrony) × two (predictability) factorial design. At the far-left and far-right of the figure, example stimulus onsets are shown for each modality as points along a horizontal timeline. In the more central parts of the figure (on shaded background), particular SOAs are plotted along the y-axis, against the successive event number within a train along the x-axis. The same 18 SOAs were used between successive events within each sensory train on every trial, drawn from a "sinusoidal" distribution of such SOAs (see central plots, on shaded background, in upper row). The exact order of these SOAs in each train of stimuli was manipulated to generate the 2 × 2 design. The temporal predictability of successive SOAs either followed a predictable/sinusoidal (top row) or unpredictable/scrambled (bottom row) SOA sequence. Orthogonal to this, the relative timing between auditory and visual trains was either synchronous (left half of figure) or asynchronous with shifted start points in the SOA sequence (right half of figure). Arrows highlight the same selected SOAs in the auditory (grey) and visual (black) sequences, to help convey how lag instead of synchrony was generated.

and fMRI measures, so the rare 5.8% of accidental synchronies in the "asynchronous" conditions evidently did not undermine our synchrony manipulation.

In 30% of trials one standard auditory event was replaced with the higher-intensity auditory target; in a separate 30% of trials one standard visual event was replaced with the higher-intensity visual target; while in the remaining 40% of trial all stimuli were standards. This produced 12 trial types in total (synchronous/asynchronous × predictable/unpredictable temporal structure × visual/auditory/no target) which were presented in a fully intermingled order. Higher-intensity targets were restricted such that they could not appear in the first six events on a given trial (thereby allowing the temporal properties of that trial to be established prior to target occurrence), nor as one of the last three events.

Examples of our stimuli from the different conditions are available for inspection online (view supplementary information), although please note that exact presentation rates when inspecting these stimuli may depend on capabilities of the computer used.

Experimental Procedures

Participants performed three functional imaging sessions of 14.4 mins, each comprising 120 trials (30 trials for each of the four audiovisual conditions produced by the 2 × 2 design for synchrony and predictability factors), plus 16 further null events (6 s), all presented in a pseudo-randomised order. Each audiovisual combination of rapid trains lasted 3 s, after which participants were given 1.5 s to make a 3-AFC button press to indicate whether a louder tone, a higher contrast flash, or no target had been present in the preceding stimuli. There was an inter-trial interval of 1.5 s. A central fixation cross (0.5°, 10.81 cd mm^{-2}) remained on the screen throughout the experimental session.

Experimental Setup

Visual and auditory stimuli were presented using Cogent v1.25 (Vision Lab, University College London, UK; http://www.vislab.ucl.ac.uk/), running in Matlab v6.5 (MathWorks, Natick, MA) on a Windows PC. Visual stimuli were back-projected onto a screen (30° × 26°) using an LCD projector (LT158; NEC, USA) with the resulting image visible to the participant inside the scanner via a mirror mounted on the MR head coil. Auditory stimuli were presented via etymotic earphones (E-A-RTONE 3A Insert Earphone, E-A-R Auditory Systems, Aearo Company, Indianapolis, USA), with external ear defenders worn to reduce background scanner noise. Participants made responses on a three-button, fiber-optic keypad with the index, middle or ring finger of their right hand, as recorded by the stimulus PC. Eye position was recorded

throughout using a long-range remote infrared video system (E5000; Applied Science Laboratories, Bedford, MA).

Behavioral Measurements

Hit-rates and false-alarm rates were each calculated separately for each target modality, as permitted by our use of separate response buttons for indicating presence of a visual or auditory target. The auditory and visual hit-rates were the proportions of correctly detected target-present trials for each respective modality. The auditory false-alarm rate was the proportion of auditory target-absent trials (i.e., trials with no target or a visual target) for which an auditory target-present response was erroneously given. The separate visual false-alarm rate was the proportion of visual target-absent trials (i.e., trials with no target or an auditory target) for which a visual target-present response was erroneously given. Hit-rates and false-alarm rates for each modality were then used to generate sensitivity (d') scores for each modality separately, via signal detection theory.

Target detection sensitivity d' ($= Z(P_{hits}) - Z(P_{false\ alarms})$) for target-present trials were calculated for each of the four audiovisual conditions, as were mean reaction times (RT), with d' or RT scores then entered into 2×2 repeated measurement ANOVAs. Trials with no recorded button response during the inter-trial interval (averaging only $3\% \pm 0.69\%$ of trials) were omitted from behavioral and fMRI analyses. All statistical analyses on behavior were performed in SPSS v16.0 (SPSS, Chicago, USA).

Scanning Protocols

A Siemens 3T Allegra MRI (Siemens, Erlangen, Germany) with head coil system was used to acquire high-resolution T_1-weighted anatomical images (176 sagittal slices, FoV = 256 mm × 240 mm FoV, 1 mm^3 voxel size); fieldmap images (double-echo FLASH, TE_1 = 10 ms, TE_2 = 12.47 ms, 3 mm × 3 mm × 2 mm resolution and 1 mm interslice gap); and T_2*-weighted echoplanar functional images for blood oxygenation level-dependent (BOLD) contrast (40 slices, 2-mm slice thickness and 1-mm gap, 3-mm resolution in plane, slice TE = 30 ms, volume TR = 2.4 s, 64 × 64 matrix). To reduce acoustic noise during scanning, we used a custom EPI sequence with a sinusoidal read out and lower slew rates [Balteau et al., 2008]. Although this sequence is slightly quieter (by 2.5 dB(A)) than the standard EPI sequence used on the Siemens Allegra at the Wellcome Trust Centre for Neuroimaging (UCL, London), the scanner sound was still audible throughout the whole session. Thus we did not use "sparse" scanning and the task was not performed in silence (although please note our use of etymotic earphones plus ear-defenders, and the constant nature of the scanner sounds regardless of experimental conditions). Three EPI sessions of 360 vol.

were collected and the first 5 vol. were discarded to allow for T_1 equilibrium effects.

fMRI Analysis

The fMRI data were submitted to statistical parametric mapping, using SPM5 software [http://www.fil.ion.ucl.ac.uk/spm; see Friston et al., 1995]. Scans from each participant were realigned using the first as a reference; unwarped incorporating fieldmap distortion information; spatially normalized into MNI standard space; resampled to 3 × 3 × 3 mm^3 voxels; then spatially smoothed with a Gaussian kernel of 8 mm FWHM, in accord with the standard SPM approach. The 12 trial types (2 levels of synchrony × 2 levels of temporal predictability × 3 target types, i.e., auditory, visual or none) were entered into an fMRI design matrix as separate regressors. These were modeled using 3-s boxcars across each of the audiovisual presentations, with a first order parametric modulator for manual reaction times added for each trial to model any brain responses relating to the speed of these motor responses. Regressors of no interest derived from the eye data were also entered: first and second order polynomial parametric modulators for mean pupil width, horizontal position, vertical position, and movement per volume were modeled, plus an additional stick-event regressor for eye blink events [see also Ruff et al., 2006, for a similar approach to eye data during fMRI]. Regressors were convolved with the canonical hemodynamic response function and its temporal derivatives in SPM5. Six further regressors derived from image realignment were entered to account for any residual head movement artifacts.

Linear compound contrast images were created to assess the main effects and interaction of the two critical audiovisual factors: synchrony and temporal predictability (collapsed across target types, which were of interest here for our behavioral measure instead). These condition-specific effects were first estimated for each participant according to the general linear model and then entered into a second-level random-effects analysis for statistical assessment across participants [Friston et al., 1999]. An initial voxel threshold was set at $t_{15} > 5$ as a prerequisite for subsequently assessing whether clusters survived correction for multiple comparisons [at $P_{FWE} < 0.05$; see Brett et al., 2003; Friston et al., 1994], as reported for our whole-brain analysis. In addition to this corrected whole-brain comparison, we also conducted region of interest (ROI) analyses for a brain area that was of particular interest a priori [in particular, for the STS region highlighted by Noesselt et al., 2007; see below]. Peak locations for all significant clusters are reported in MNI space.

An a priori ROI was preselected at a site in left posterior STS ($x = -54$, $y = -50$, $z = 8$). This site was previously identified in the Noesselt et al. [2007] study as corresponding to a multisensory region preferentially activated
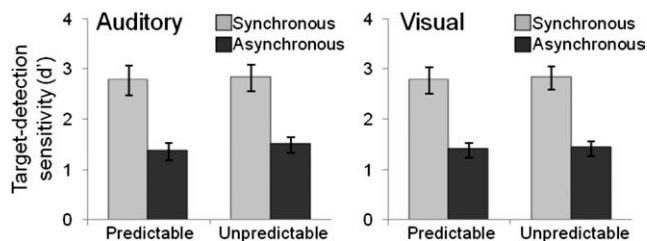
Figure 2.
Target-detection sensitivity (*d'*). Auditory (left bar-graph) and visual (right bar-graph) target-detection sensitivity (*d'*) were both enhanced by synchronous (grey bars) compared to asynchronous (black bars) audiovisual presentations ($P < 0.001$), regardless of whether the stimulus trains were predictable or unpredictable. Group means plotted (± 1 s.e.m.).

during synchronous rather than asynchronous audiovisual trains. On the basis of their previous study, we predicted enhanced BOLD signal in this region for synchronous versus asynchronous audiovisual presentations. We could also test here whether this multisensory STS region would also show any influence of temporal predictability (that might potentially modulate the impact of synchrony); and any relation to behavioral impacts upon higher-intensity target detection. To interrogate the left STS ROI as previously identified by Noesselt et al. [2007], we extracted parameter estimate beta values for each voxel within an 8-mm sphere centered at the predefined Noesselt et al. coordinates of $x = -54$, $y = -50$, $z = 8$, then averaged across voxels within that sphere using the MarsBaR toolbox [Brett et al., 2002]. The resulting values for each participant were then entered into ROI contrasts for the effects of interest.

We also implemented a robust regression analysis (MATLAB robustfit function, default bisquare option) to test for any relation between the mean percentage BOLD signal change (extracted via the MarsBar toolbox) in the ROI as a function of experimental condition, with the observed subject-by-subject behavioral change in auditory *d'* or visual *d'* for target detection in the corresponding conditions. Specifically, we compared the change in d' for each target modality (scored separately) during synchronous versus asynchronous presentations with the change in BOLD percentage signal for the same contrast. Note that by applying this test for brain–behavior relations to an independently predefined ROI, we were able to avoid the selection biases and circularity that can otherwise arise due to potential "double-dipping" [Kriegeskorte et al., 2009]. Note also that by using the robustfit function, we could guard statistically against any such brain–behavior relations being driven primarily by unrepresentative outliers.

For completeness at the request of a reviewer, we also implemented analogous brain–behavior robust regression analyses for all clusters activated more by synchronous

than asynchronous conditions (see below). Although *post-hoc*, these brain–behavior regressions were again applied to clusters that were initially defined separately from any consideration of behavior, to avoid circularity. Because one such cluster encompassed both the left putamen and thalamus, we used anatomical masks [MarsBar AAL ROI library; Tzourio-Mazoyer et al., 2002] to separate this cluster into its constituent anatomical parts.

RESULTS

Audiovisual Synchrony Enhances Behavioral Detection Sensitivity (*d'*) for Higher-Intensity Targets Among Standards in Either Modality

Mean behavioral target-detection sensitivity (*d'*) for each modality is plotted in Figure 2 as a function of audiovisual synchrony and predictability of the temporal trains. Target-detection sensitivity was enhanced by audiovisual synchrony, but predictability had no impact. A three-way repeated-measure ANOVA (synchrony × predictability × target modality) confirmed a significant main effect of audiovisual synchrony on *d'* ($F_{(1,15)} = 102.8$, $P < 0.001$). Audiovisual trains that were synchronous led to enhanced *d'* (mean ± S.E.M, 2.76 ± 0.23) compared to asynchronous conditions (1.41 ± 0.13). Note that this is by definition a multisensory effect, since only the relationship between modalities varied as a function of synchrony versus asynchrony; the nature of events within each single modality, when considered alone, was fully conserved regardless of synchrony. This impact of synchrony did not interact with target modality ($F_{(1,15)} = 0.125$, $P = 0.73$) nor predictability ($F_{(1,15)} = 0.159$, $P = 0.70$) and there were no other significant terms in the three-way ANOVA.

Reaction times for target-present trials were also facilitated for synchronous (mean 665 ± 40 ms) versus asynchronous (735 ± 45 ms) conditions, leading again to a main effect of synchrony ($F_{(1,15)} = 21.506$, $P < 0.001$), but no other significant terms in a comparable three-way ANOVA on the RT data (Table I). Hence the RT pattern agrees with the *d'* pattern for behavioral results. For both measures, performance was enhanced by audiovisual synchrony (even though this synchrony in itself gave no information about whether a particular event was a target or nontarget); but

TABLE I. Reaction times for target-present trial judgments

	Auditory target trials	Visual target trials
Synchronous predictable	676 (±46) ms	640 (±39) ms
Synchronous unpredictable	687 (±48) ms	657 (±39) ms
Asynchronous predictable	720 (±46) ms	729 (±42) ms
Asynchronous unpredictable	765 (±55) ms	727 (±44) ms

Groups means (±1 s.e.m.) reported.
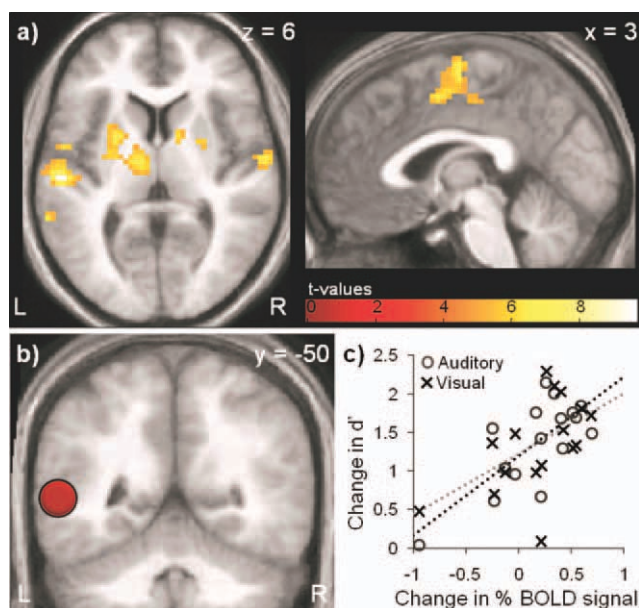
Figure 3.

Synchrony effect on BOLD signal in multisensory and auditory cortex. (**a**) Activations in multisensory areas (left posterior STS), auditory cortex (bilateral STG), thalamus and regions previously implicated in timing processes (supplementary motor area, putamen) all showed significantly higher BOLD signals for synchronous versus asynchronous conditions (voxel $t > 5$ and cluster $P_{FWE} < 0.05$), regardless of predictability; see Table II. (**b**) A synchronous > asynchronous effect was also found in a predefined left STS ROI (8 mm sphere centered at $[-54\ -50\ 8]$ co-ordinates taken from Noesselt et al. [2007], as indicated schematically here by red circle. (**c**) In a robust-regression analysis (see main text), the impact of synchrony > asynchrony (the "change" plotted corresponds to the difference in this subtraction) for percent BOLD signal in each participant was found to be positively related to their synchrony > asynchrony behavioral effect (the "change" plotted again corresponds to this subtraction) for both auditory ($P = 0.002$; grey circles) and visual ($P = 0.028$; black crosses) target detection sensitivity, d'. Lines shown are from the Robustfit regressions for each modality.

there was no impact of the temporal predictability of the stimulus trains on behavioral target detection.

fMRI Data: Audiovisual Synchrony Enhances BOLD in STS and Auditory Cortex Plus a Wider Network

Whole-brain analysis revealed significant enhancement of BOLD signal in a network of brain regions due to audiovisual synchrony (see Fig. 3a and Table II). A main effect of synchrony > asynchrony was found for the left posterior STS as predicted, plus further areas [as was also the case in Noesselt et al., 2007] known to be involved in auditory

processing (bilateral superior temporal gyri, including Heschl's gyrus and planum temporale). Several further areas (SMA; left precentral and postcentral gyri; bilateral putaman and thalamus) were also activated. While some of these additional areas would traditionally be associated with motor-related processing, several of these regions have been activated in recent studies on timing processes [e.g., Bengtsson et al., 2009; Grahn and Brett, 2007; Grahn and Rowe, 2009] which may explain their sensitivity to synchronous timing here. No brain areas were more active for the asynchronous than synchronous conditions.

As mentioned earlier, in addition to the whole-brain analysis we also focused on a left posterior STS ROI that had been selected *a priori*, based on the results from the related previous study of Noesselt et al. [2007]. They had particularly emphasized their finding of higher BOLD for synchronous than unrelated audiovisual trains in posterior left STS, peaking at $x = -54$, $y = -50$, $z = 8$ in their study. An 8-mm spherical ROI centered on these coordinates (see schematic in Fig. 3b) also showed a main effect of synchrony > asynchrony as reported above ($t_{15} = 2.549$, $P_{(1-tail)} = 0.01$), that did not interact with predictability. This STS ROI was not the sole region to show the main effect of audiovisual synchrony here (see Fig. 3a and Table II), as had also been the case in Noesselt et al. [2007]; see their Table II. But the STS ROI was nevertheless of major *a priori* interest here, due to Noesselt et al.'s emphasis upon it and also the wider interest in STS for audiovisual multisensory fMRI studies [see also Beauchamp et al., 2004a,b; Calvert et al., 2001; Macaluso et al., 2004; Stevenson and James, 2009; Stevenson et al., 2010; Werner and Noppeney, 2010]. It is thus noteworthy that the same STS ROI as in Noesselt et al. again showed enhanced BOLD due to audiovisual synchrony here. Moreover a novel finding was that this impact of audiovisual synchrony on the STS ROI was found regardless of the new factor of temporal predictability, which had no significant influence on this ROI.

To address a reviewer request, we seeded a simple "effective connectivity" analysis in the STS ROI, testing for condition-dependent changes in residual covariation with other remote brain areas, as a function of condition. This "psychophysiological interaction" (PPI) analysis [Friston et al., 1997] revealed no significant remote covariations with STS as a function of condition. We next turn to the possible relation of the audiovisual synchrony impact on BOLD in the STS ROI, that was unaffected by predictability, to the corresponding impact of audiovisual synchrony on behavioral target detection (cf. Fig. 2), which likewise was not modulated by predictability.

Brain–Behavior Relation Within the STS ROI for the Effect of Audiovisual Synchrony

We used the same independently-defined STS ROI to assess any relation between the BOLD effect due to synchrony and the behavioral enhancement of target-detection

TABLE II. Brain clusters with higher BOLD signal for synchronous versus asynchronous audiovisual presentations at corrected significance

		Size (voxels)	Cluster P_{FWE}	Peak z-score	MNI coordinates		
					x	y	z
-	supplementary motor area	145	<0.001	4.61	0	−12	66
L	superior temporal gyrus	139	<0.001	5.37	−54	−21	6
L	superior temporal sulcus	12	0.024	4.48	−63	−45	6
L	postcentral gyrus	23	0.002	4.11	−39	−33	51
L	putamen/thalamus	278	<0.001	5.10	−24	−9	−9
R	superior temporal gyrus	24	0.002	4.44	63	−9	6
R	insula	15	0.011	4.72	39	−3	−6
R	putamen	32	<0.001	4.14	27	0	6
R	caudate	12	0.024	4.21	12	3	6

Peak voxel co-ordinates (MNI space) and statistical values ($t > 5$) are listed for significant clusters ($P_{FWE} < 0.05$).

sensitivity (d') for the synchronous versus asynchronous conditions. Since the ROI was predefined based on Noesselt et al. [2007], this circumvented any "double-dipping" problems that can otherwise arise in some searches for brain–behavior relations [Kriegeskorte et al., 2009]. We implemented a robust regression which revealed that across participants ($n = 16$), the increase in percentage BOLD signal for the STS ROI in synchronous minus asynchronous conditions was significantly related to the participant-by-participant increase in detection sensitivity (d') for the synchronous conditions, for both auditory targets ($y = 1.027x + 1.187$; $t_{15} = 3.86$, $P = 0.002$) and visual targets ($y = 0.784x + 1.213$; $t_{15} = 2.45$, $P = 0.028$) targets (see Fig. 3c). This indicates that participants with a greater increase in BOLD signal for synchronous audiovisual presentations, within the independently-defined STS ROI, also tended to have a greater perceptual increase for target detection in both senses during audiovisual synchrony.

For completeness, a reviewer asked that we perform similar brain–behavior regressions for all of the regions that had been significantly activated by audiovisual synchrony versus asynchrony in our whole-brain analysis (cf. Table II). The outcome of this *post hoc* further analysis (cf. our *a priori* focus on the STS ROI) is shown in Supporting Information Table SI. In brief, some further brain–behavior relations were found in this way for bilateral putamen, extending into left thalamus. Separation of the left putamen/thalamus cluster by anatomical masks further revealed that BOLD signal change in the putamen appeared positively related to both visual and auditory d' measures; while the thalamic portion appeared positively related only to auditory d' [see also Noesselt et al., 2010]. The reviewer also suggested that we conduct a further whole-brain test for any brain–behavior relations to determine if any further regions (beyond those in Table II and Supporting Information Table SI) might show such relations; none were found (neither at our corrected thresholds, nor at a less stringent criterion of $P < 0.001$ uncorrected). Likewise a whole-brain search for any regions showing significant differential

BOLD response for correct versus incorrect trials found no such areas other than left pallidum (see Supporting Information, including Table SII) that we shall not discuss further as this was unexpected.

Synchrony Effect on BOLD Modulated by Temporal Predictability Only in Right Inferior Parietal Cortex

We also examined whole-brain SPMs for any significant interaction between synchrony and temporal predictability in BOLD signals. Testing for stronger effects of synchrony > asynchrony during predictable than unpredictable trains revealed a significant interaction only in right inferior parietal cortex (peaking at $x = 57$, $y = −48$, $z = 42$; peak z-score $= 4.11$; cluster $P_{FWE} = 0.028$) see Figure 4. Pairwise t tests confirmed that BOLD signals in this cluster showed a
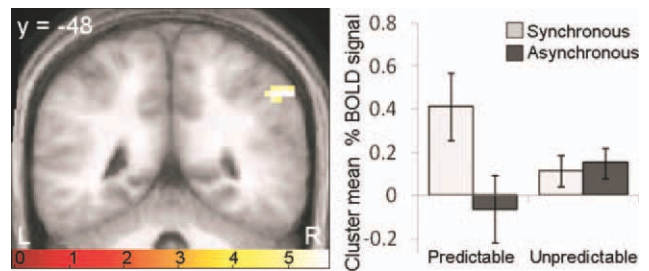
Figure 4.
Temporal predictability modulates the audiovisual synchrony effect in rIPL. Difference between synchronous and asynchronous conditions was modulated by temporal predictability for the BOLD signal in right inferior parietal lobule (rIPL) (interaction SPM shown here thresholded at voxel $t > 5$ and cluster $P_{FWE} < 0.05$). The plot of percentage BOLD signal for each audiovisual condition illustrates the nature of this interaction; see main text. Group means plotted (± 1 s.e.d. for synchrony versus asynchrony differences).

significant enhancement for synchronous compared to asynchronous conditions only for trials with predictable timing ($t_{15} = 3.0$, $P_{(2\text{-tail})} = 0.009$), but not for trials with unpredictable timing ($t_{15} = -0.5$, $P_{(2\text{-tail})} = 0.489$, n.s.); see corresponding plot in Figure 4. None of the regions which had shown a main effect of audiovisual synchrony (Fig. 3 and Table II) showed such an interaction; nor did the STS ROI.

Furthermore there were no significant clusters (and no impact on the STS ROI) for the reverse interaction that we had motivated as a theoretical possibility in our Introduction; namely stronger synchrony effects for unpredictable than predictable trains. Finally no brain regions showed any significant main effects of predictability.

DISCUSSION

We tested the impact of audiovisual synchrony between temporally unpredictable or predictable stimulus trains, both for performance in a target-detection task performed on both modalities; and also for fMRI activations examined with corrected whole-brain analysis and for a multisensory STS ROI motivated *a priori* by other recent fMRI work [Noesselt et al., 2007; see also Beauchamp et al., 2004a,b; Bischoff et al., 2007; Calvert et al., 2001; Hein et al., 2007; Macaluso et al., 2004; Meienbrock et al., 2007; Stevenson and James 2009; Stevenson et al., 2010; Werner and Noppeney, 2010]. The Noesselt et al. [2007] study provided one close precedent for the current work. But unlike here they did not impose a behavioral target-detection task for the stimulus trains in the two modalities. Nor did they vary the temporal predictability of these trains (they had used only unpredictable, temporally erratic streams). Hence, as explained in our Introduction, their finding of higher STS activation during audiovisual synchrony might have been specific to their very unpredictable context, for which only synchrony can provide a simplifying "generative-model" of the otherwise erratic sensory inputs.

Here we found higher BOLD during audiovisual synchrony than asynchrony, within a network encompassing auditory cortex, STS, plus other regions previously associated with timing functions; see Figure 3a and Table II. Moreover we also found this effect of audiovisual synchrony specifically for the STS ROI centered at the left STS region finding emphasized by Noesselt et al. [2007]. Importantly, none of these regions that showed a main effect of audiovisual synchrony were affected by our new predictability factor, including the STS ROI. Furthermore there was no significant interaction anywhere in the brain of the specific form that was theoretically motivated in our Introduction (namely a potentially larger impact of audiovisual synchrony for unpredictable trains in particular). The present results thus suggest that the impact of audiovisual synchrony on these regions (including the STS ROI) arises in the same bottom-up manner, regardless of predictable or unpredictable temporal context. Put another way, the central finding in Noesselt et al. [2007]—namely increased activation of STS and further regions due to audiovisual synchrony—evidently does not depend on the erratic, temporally unpredictable nature of the stimuli that they had used. Here we find this activation pattern regardless of whether the stimulus trains are unpredictable or predictable.

Turning to behavior, a further advance on previous work is that here we were able to document an impact of audiovisual synchrony for a target-detection task performed on the same trains that led to the BOLD effects. Specifically, target-detection sensitivity (d') for higher-intensity targets differed for both modalities between the synchronous than asynchronous conditions, with objectively better performance in the synchronous than asynchronous presentations (see Fig. 2). Analogously to the fMRI results mentioned above, this impact of audiovisual synchrony on behavior arose regardless of whether the trains were predictable or unpredictable. This behavioral effect of audiovisual synchrony is of interest in its own right. It represents a non-trivial multisensory finding, since the occurrence of a higher-intensity target in one modality was never signaled by the nature of events in the other modality.

Moreover here we were able to link these significant behavioral effects on auditory and visual d' due to audiovisual (a) synchrony to the corresponding BOLD effects on the independently pre-defined STS ROI. Participants showing larger BOLD effects of synchrony in STS also tended to show larger benefits in performance, as confirmed for both auditory and visual performance with robust regression of the BOLD data against the behavioral d' scores (see scatter-plot in Fig. 3c). This suggests that the responsivity of STS to audiovisual synchrony does relate to multisensory benefits in performance for each modality. A similar brain–behavior relation was also observed post hoc for putamen/thalamus, indicating that STS may be just one part of a wider network.

The participant-by-participant brain–behavior relation that we observed for the STS ROI is reminiscent in some respects of a recent study by Werner and Noppeney [2010]. Using degraded videos and auditory clips of tools or musical instruments, they reported a positive relation between the multisensory impacts upon activity in a similar region of STS and upon behavioral task sensitivity (d'), when collapsing across tasks of object classification or detection of an embedded tone and/or flash target. Their study compared responses to multisensory object stimuli presented to both senses (congruent and synchronous) against a combination of unisensory responses to stimuli in which the object was presented to one sense only, while white noise was presented concurrently to the other sense (incongruent). Although they presented auditory and visual stimuli for all conditions, their target object was presented to both senses in the "congruent" condition (multisensory targets) but to only one sense in the "incongruent" condition (unisensory targets). Therefore their

congruent multisensory condition arguably provided more information about targets. This contrasts in some respects with our study, where each target event (if present) was only defined in one particular modality for all our conditions, yet we could still measure the beneficial impact of audiovisual synchrony upon detection of targets in either modality. Thus here we were able to show that temporally synchronous audiovisual presentations can enhance target detection even when synchrony (versus asynchrony) provides no information about whether a particular event was a target stimulus, or instead a standard stimulus. Our specific findings thus join a wider literature documenting multisensory effects that can arise even when a second modality provides no objective information about the target-defining property in a first modality [see Noesselt and Driver, 2008].

Here our critical comparisons were made between synchronous and asynchronous multisensory conditions. As noted by a reviewer, it might be interesting in future extensions of the current work to include unisensory baselines also. Those might in principle allow tests for whether audiovisual synchrony enhances performance and related brain activations relative to unisensory baselines; while audiovisual asynchrony might impair performance and related brain activations relative to such baselines. This was not implemented here, however, because any such unisensory baselines would require only one modality to be attended for the target-detection task. Here instead both modalities always had to be attended in all our conditions, making our contrasts well-controlled in that respect.

Another potentially interesting future direction, suggested by another reviewer, might be to examine the impact of synchrony and temporal predictability for more naturalistic stimuli than the tones and flashes that were used here for simplicity and experimental control. Stevenson and James [2009] observed that different types of naturalistic audiovisual stimuli (e.g., speech, objects) can produce multisensory effects at somewhat different locations along the STS [see also Calvert et al., 2001; Macaluso et al., 2004]. Extending our own paradigm to more naturalistic stimuli could be a useful step, although here we deliberately avoided stimuli with semantic associations in order to isolate any effects due to audiovisual synchrony and/or temporal predictability per se.

Moving beyond STS, we also found that auditory cortex (superior temporal gyri) showed enhanced BOLD for audiovisual synchrony. Noesselt et al. [2007] had also found in the strongest effect of audiovisual synchrony within the temporal lobe to arise in the STG (see their Table II). This may accord with other demonstrations that some parts of auditory cortex can receive convergent input from other senses [e.g., see Ghazanfar et al., 2005; Lakatos et al., 2007; see also Bizley et al., 2007; Kayser et al., 2007, 2008; Schroeder and Foxe, 2005] via direct or indirect anatomical connections [see Falchier et al., 2002, 2010]. We did not find any impact of synchrony upon visual cortex,

unlike Noesselt et al. [2007]. Since null outcomes in fMRI have to be treated with caution, we will not make much of this, except to note that a different outcome might have been found if using visual stimuli more similar to those used by Noesselt et al. But the this result does not undermine our positive results for STG and STS; nor the behavioral findings of enhanced target-detection d' due to audiovisual synchrony; nor the brain–behavioral relation we found for the STS ROI in particular, which was not present for other cortical regions.

The lack of audiovisual synchrony effects in visual cortex here (e.g., for the calcarine sulcus) also appears somewhat different to a recent study by Lewis and Noppeney [2010], but their study differed from ours in many respects. They used a visual rotational-motion task, finding performance benefits and increases in visual cortex BOLD when auditory clicks were made synchronous (versus asynchronous) with the visual rotations that had to be judged. Moreover, activity in V5/hMT+ related to the subject-by-subject benefits in visual discrimination for rotational motion that they found due to synchronous (vs. asynchronous) clicks. But their design was not closely comparable to ours. Here both modalities were always task-relevant, whereas only vision was judged in their study; targets were defined by intensity-differences here, rather than higher-level properties such as motion; and audiovisual synchrony here did not provide any information about which visual time-points provided target information. Finally, visual targets were not embedded among visual noise-elements here, unlike Lewis and Noppeney [2010].

In addition to affecting STS and auditory cortex, synchronous audiovisual presentations here also led to increased BOLD in the thalamus, some parts of which may conceivably act as a multisensory relay centre [for reviews see Cappe et al., 2009; Smiley and Falchier, 2009; see also Noesselt et al., 2010]. The SMA and putamen were also affected; these areas are thought to be involved in temporal analysis and rhythmic prediction [Bengtsson et al., 2009]. Greater activity [Grahn and Brett, 2007] or functional connectivity between these regions and auditory cortex [Grahn and Rowe, 2009] has previously been demonstrated for some structured temporal sequences (e.g., regular isometric beats). But we found no impact of temporal predictability versus unpredictability with the stimuli used here. We note that unlike past work on regular beats, our stimulus trains lasted only three seconds (see Fig. 1), had a rapid rate of events (6 Hz on average), and even our "predictable" conditions never presented a regular isometric beat (instead following a sinusoidal distribution of intervals between successive events). On the other hand, the brain was evidently sensitive to the particular form of temporal predictability we used here, as indicated by the final result we shall discuss.

Specifically, right inferior parietal cortex showed an impact of audiovisual synchrony only for the predictable (sinusoidal) temporal patterns. BOLD signal here was highest in the predictable synchronous condition, which is

the most constrained situation and thus might be considered as lowest in entropy or "free energy" [Friston, 2010]. This finding confirms that even though our predictability manipulation did not affect the target-detection task or the synchrony effects in STS and STG, this manipulation was nevertheless effective and the brain could detect the temporal predictability when present. Moreover, it is noteworthy to identify a brain region, here in right parietal cortex, which evidently combines information about the predictability of each temporal pattern with the synchronous relation to the other modality. Context-specific modulation of audiovisual synchrony effects on right inferior parietal cortex activity has previously been shown for synchrony effects with audiovisual speech stimuli, as a function of common or discrepant spatial location [Macaluso et al., 2004]. Accordingly we speculate that this region may be involved detecting non-accidental relations between multiple properties of audiovisual stimuli. But we note also that here behavioral target detection was not affected by the particular interaction between predictability and synchrony that arose for right inferior parietal cortex, only by audiovisual synchrony (as for STS and STG). Moreover the inferior parietal BOLD effects did not relate systematically to participant-by-participant behavioral patterns, unlike the STS. Hence our conclusions focus primarily on the pattern found for STS ROI, which was the area of main *a priori* interest in any case, given prior work.

CONCLUSIONS

We found that audiovisual synchrony enhanced the BOLD response in posterior STS (plus a wider network of regions including STG and subcortical structures), regardless of whether the stimulus trains were predictable or unpredictable. Likewise target-detection sensitivity (d') for higher-intensity targets was enhanced by audiovisual synchrony (again regardless of temporal predictability of the stimulus trains), even though each target was defined by intensity within only one modality. The effect of audiovisual synchrony on BOLD in a left STS ROI [predefined by the separate data of Noesselt et al., 2007] related systematically to the participant-by-participant behavioral d' effect for both auditory and visual targets. Several other regions (auditory cortex, SMA, putamen, thalamus) also showed stronger BOLD signals during auditory-visual synchrony; while the right inferior parietal cortex was unique in showing an impact of audiovisual synchrony only for predictable temporal patterns. Our results indicate that STS is sensitive to audiovisual synchrony, regardless of temporal predictability, and may mediate the impact of audiovisual synchrony on behavioral sensory performance.

ACKNOWLEDGMENTS

REFERENCES

Balteau E, Josephs O, Ruff C, Weiskopf N (2008): Quiet EPI Sequence With Sinusoidal Readout for fMRI Studies. Presented at the 25th European Society for Magnetic Resonance in Medicine and Biology Conference; 2008 Oct 2–4; Valencia, Spain.

Beauchamp MS (2005): See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex. Curr Opin Neurobiol 15:145–153.

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

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

Bengtsson SL, Ullen F, Ehrsson HH, Hashimoto T, Kito T, Naito E, Forssberg H, Sadato N (2009): Listening to rhythms activates motor and premotor cortices. Cortex 45:62–71.

Bischoff M, Walter B, Blecker CR, Morgen K, Vaitl D, Sammer G (2007): Utilizing the ventriloquism-effect to investigate audiovisual binding. Neuropsychologia 45:578–586.

Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007): Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb Cortex 17:2172–2189.

Brett M, Anton J-L, Valabregue R, Poline J-B (2002): Region of Interest Analysis Using an SPM Toolbox. Presented at the 8th International Conference on Functional Mapping of the Human Brain; 2002 June 2–6; Sendai, Japan.

Brett M, Penny W, Kiebel,S (2003):An introduction to random field theory. In: Frackowiak RSJ,Friston KJ,Frith CD,Dolan RJ,Price CJ,Zeki S,Ashburner J,Penny W, editors. Human Brain Function, 2nd ed.Oxford, UK:Elsevier. p877.

Bushara KO, Grafman J, Hallett M (2001): Neural correlates of auditory-visual stimulus onset asynchrony detection. J Neurosci 21:300–304.

Calvert CA, Spence C, Stein BE, editors (2004): The Handbook of Multisensory Processing. Cambridge, MA: 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 audiovisual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage 14:427–438.

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

Dayan P, Hinton GE, Neal RM, Zemel RS (1995): The Helmholtz machine. Neural Comput 7:889–904.

den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009): A dual role for prediction error in associative learning. Cereb Cortex 19:1175–1185.

Doehrmann O, Naumer MJ (2008): Semantics and the multisensory brain: How meaning modulates processes of audiovisual 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.

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

Falchier A, Schroeder CE, Hackett TA, Lakatos P, Nascimento-Silva S, Ulbert I, Karmos G, Smiley JF (2010): Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. Cereb Cortex 20:1529–1538.

Frassinetti F, Bolognini N, Ladavas E (2002): Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp Brain Res 147:332–343.

Friston KJ (2005): A theory of cortical response. Philos Trans R Soc Lond B Biol Sci 360:815–836.

Friston KJ (2010): The free-energy principle: A unified brain theory? Nat Rev Neurosci 11:127–138.

Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans JC (1994): Assessing the significance of focal activations using their spatial extent. Hum Brain Mapp 1:214–220.

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

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997): Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229.

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

Furl N, Kumar S, Alter K, Durrant S, Shawe-Taylor J, Griffiths TD (2011): Neural prediction of higher-order auditory sequence statistics. Neuroimage 54:2267–2277.

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

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

Grahn JA, Brett M (2007): Rhythm and beat perception in motor areas of the brain. J Cogn Neurosci 19:893–906.

Grahn JA, Rowe JB (2009): Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat perception. J Neurosci 29:7540–7548.

Hein G, Doehrmann O, Muller 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.

Helmholtz H (1860): In: Southall JPD, editor. Helmholtz's Treatise on Physiological Optics (1962), Vol.3. New York: Dover (English translation).

Hertz U, Amedi A (2010): Disentangling unisensory and multisensory components in audiovisual integration using a novel multifrequency fMRI spectral analysis. Neuroimage 52:617–632.

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 258:80–88.

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009): Circular analysis in systems neuroscience: The dangers of double dipping. Nat Neurosci 12:535–540.

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.

Lewis R, Noppeney U (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. J Neurosci 30:12329–12339.

Lovelace CT, Stein BE, Wallace MT (2003): An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection. Brain Res Cogn Brain Res 17:447–453.

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

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

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.

McIntosh AR, Cabeza RE, Lobaugh NJ (1998): Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. J Neurophysiol 80:2790–2796.

Meienbrock A, Naumer MJ, Doehrmann O, Singer W, Muckli L (2007): Retinotopic effects during spatial audiovisual integration. Neuropsychologia 45:531–539.

Meredith MA, Nemitz JW, Stein BE (1987): Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci 7:3215–3229.

Mumford D (1992): On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol Cybern 66:241–251.

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

Noesselt T, Tyll S, Boehler CN, Budingder E, Hienze H-J, Driver J (2010): Sound-induced enhancement of low-intensity vision: Multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. J Neurosci 30(21):13609–13623.

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.

Odgaard EC, Arieh Y, Marks LE (2003): Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. Percept Psychophys 65:123–132.

Odgaard EC, Arieh Y, Marks LE (2004): Brighter noise: Sensory enhancement of perceived loudness by concurrent visual stimulation. Cogn Affect Behav Neurosci 4:127–132.

Overath T, Cusack R, Kumar S, von Kriegstein K, Warren JD, Grube M, Carlyon RP, Griffiths TD (2007): An information theoretic characterization of auditory encoding. PLOS Biol 5:e288.

Rao RP (1999): An optimal estimation approach to visual perception and learning. Vis Res 39:1963–1989.

Ruff CC, Blankenburg F, Bjoertomt O, Bestmann S, Freeman E, Haynes J-D, Rees G, Josephs O, Deichmann R, Driver J (2006): Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. Curr Biol 16:1479–1488.

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

Smiley JF, Falchier A (2009): Multisensory connections of monkey auditory cerebral cortex. Hear Res 258:37–46.

Spence C, Driver J, editors (2004): Crossmodal Space and Cross-modal Attention. Oxford, UK: Oxford University Press.

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

Stein BE, Wallace MT (1996): Comparisons of cross-modality integration in midbrain and cortex. Prog Brain Res 112:289–299.

Stein BE, London N, Wilkinson LK, Price DD (1996): Enhancement of perceived visual intensity by auditory stimuli: A psycho-physical analysis. J Cogn Neurosci 8:497–506.

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

Stevenson RA, Altieri NA, Kim S, Pisoni DB, James TW (2010): Neural processing of asynchronous audiovisual speech perception. Neuroimage 49:3308–3318.

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

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

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

Werner S, Noppeney U (2010): Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cereb Cortex 20:1829–1842.