Spatial Orienting in Complex Audiovisual Environments

Davide Nardo,[1]* Valerio Santangelo,[1,2] and Emiliano Macaluso[1]

[1]*Neuroimaging Laboratory, Santa Lucia Foundation, Rome, Italy*
[2]*Department of Human and Educational Sciences, University of Perugia, Perugia, Italy*

Abstract: Previous studies on crossmodal spatial orienting typically used simple and stereotyped stimuli in the absence of any meaningful context. This study combined computational models, behavioural measures and functional magnetic resonance imaging to investigate audiovisual spatial interactions in naturalistic settings. We created short videos portraying everyday life situations that included a lateralised visual event and a co-occurring sound, either on the same or on the opposite side of space. Subjects viewed the videos with or without eye-movements allowed (overt or covert orienting). For each video, visual and auditory saliency maps were used to index the strength of stimulus-driven signals, and eye-movements were used as a measure of the efficacy of the audiovisual events for spatial orienting. Results showed that visual salience modulated activity in higher-order visual areas, whereas auditory salience modulated activity in the superior temporal cortex. Auditory salience modulated activity also in the posterior parietal cortex, but only when audiovisual stimuli occurred on the same side of space (multisensory spatial congruence). Orienting efficacy affected activity in the visual cortex, within the same regions modulated by visual salience. These patterns of activation were comparable in overt and covert orienting conditions. Our results demonstrate that, during viewing of complex multisensory stimuli, activity in sensory areas reflects both stimulus-driven signals and their efficacy for spatial orienting; and that the posterior parietal cortex combines spatial information about the visual and the auditory modality. *Hum Brain Mapp 35:1597–1614, 2014.* © 2013 Wiley Periodicals, Inc.

Key words: attention; space; visual; auditory; multisensory; eye movements; saliency; orienting; ecological; posterior parietal cortex

INTRODUCTION

In everyday life, the brain has to process a multitude of sensory streams that continuously stimulate our senses. Signals in different modalities can jointly influence many different aspects of perception (e.g., speech comprehension, McGurk and MacDonald, 1976; object recognition, Giard and Peronnet, 1999; and spatial orienting, Spence and Driver, 1997). In the spatial domain, previous studies have highlighted that concurrent stimulation in different modalities but at one single location can improve whole-body orienting responses in animals [Stein et al., 1989] and speed up eye-movements towards the stimuli-location in humans [Arndt and Colonius, 2003; Corneil and Munoz, 1996; Corneil et al., 2002]. Crossmodal spatial interactions do not only influence overt orienting behaviour, but can also affect sensory processing. When volunteers are asked to maintain central fixation, presenting a task-irrelevant 'cue' in one modality can facilitate the detection/discrimination of a subsequent 'target' in a different modality, selectively when cue and target are presented on the same

Contract grant sponsors: The European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC grant agreement 242809; The Italian Ministry of Health.

*Correspondence to: Davide Nardo, Neuroimaging Laboratory, Santa Lucia Foundation, Via Ardeatina 306, Rome 00179, Italy. E-mail: davidenardo@gmail.com

Received for publication 9 November 2012; Revised 22 January 2013; Accepted 7 February 2013

DOI: 10.1002/hbm.22276

Published online 24 April 2013 in Wiley Online Library (wileyonlinelibrary.com).

side of space [McDonald et al., 2000; Spence et al., 1998]. These crossmodal and spatially specific effects are though to arise because of supramodal mechanisms for the allocation of spatial attention [Farah et al., 1989; Macaluso, 2010].

Previous research on crossmodal spatial interactions has employed simple and stereotyped stimuli (e.g., flashes of light and bursts of white noise), and little is known about behavioural and neurophysiological aspects of spatial orienting in complex multisensory situations. The investigation of crossmodal interactions with complex stimuli is most relevant with respect to the possible interplay between attention and multisensory processing [Koelewijn et al., 2010; Talsma et al., 2010]. Complex, 'life-like' situations entail the presence of multiple co-occurring stimuli that compete for attentional resources, and include some incertitude about the spatial correspondence between events in different modalities. A sound may originate from one or another object in a complex visual scene, which implies that further selection mechanisms are needed to associate the sound with the corresponding visual object/event. This is in striking contrast with traditional experimental paradigms typically comparing trials with one visual and one auditory stimulus, which can be unambiguously classified as being either at the same or at the different locations (e.g., valid vs. invalid trials in crossmodal spatial cueing paradigms; see Macaluso et al., 2000; McDonald and Ward, 2000; Spence and Driver, 1997).

Here, we investigated audiovisual (AV) spatial interactions in naturalistic settings, using short AV videos portraying everyday life situations. Computational models of visual and auditory salience were used to index the stimulus-driven, attention-capturing strength of visual (VEs) and auditory events (AEs) in each video. Visual saliency (Vsal) maps [Itti and Koch, 2001; Itti et al., 1998] have been previously used to study spatial orienting within naturalistic environments (e.g., pictures, Elazary and Itti, 2008; and videos, Carmi and Itti, 2006). In the auditory modality, Kayser and colleagues (2005) characterised auditory salience by using a computational architecture analogous to the visual salience model [see also Kalinli and Narayanan, 2009]. It should be noted that while the Vsal maps identify specific locations that are more 'attention-grabbing' than others, auditory saliency (Asal) maps do not provide any such spatial information (indexing the strength of auditory signals in the time–frequency domain, instead). Given our specific interest in the spatial interactions between the two modalities, here we controlled the position of the auditory stimuli—and therefore the AV spatial relationship—by presenting sounds unilaterally either from the left-hand side or from the right-hand side of space.

From a behavioural point of view, we expected that salient visual signals would attract subject's gaze/attention [Elazary and Itti, 2008; Itti and Koch, 2001], and that auditory signals in the same hemifield would further strengthen any such effect on orienting behaviour (Frens et al., 1995; see also Amlôt et al., 2003; for related results in the visuotactile domain). A central question here was how the saliency of signals in each modality would influence any spatial interaction between the two modalities. On the one hand, it can be hypothesised that the stronger the signals in vision and audition on the same side, the greater the tendency to spatially orient towards that side [Onat et al., 2007]. By contrast, crossmodal interactions may be inversely related to stimulus strength (e.g., 'inverse effectiveness', Stein and Meredith, 1993; 'optimal cue weighting', Ernst and Banks, 2002), with larger crossmodal interactions expected when the unimodal signals are weak and/or unreliable (Lewis and Noppeney, 2010; Helbig et al., 2012; for recent studies considering both behaviour and functional magnetic resonance imaging [fMRI]). In this framework, our current manipulation of visual and auditory salience may lead to larger crossmodal spatial interactions when one (or both) modality provides weak unimodal spatial cues; for example, the spatial congruence between visual and auditory stimuli may affect orienting behaviour only when the salience of the visual event (VE) is low.

From the neuroimaging perspective, the most likely candidates to process AV spatial signals in complex environments include association areas in frontoparietal cortex, as well as sensory-specific regions (e.g., the visual occipital cortex) that have been identified in previous studies on AV spatial interactions [Meienbrock et al., 2007; Santangelo et al., 2009]. The dorsal frontoparietal attention system has been traditionally linked with the control of spatial orienting behaviour [Andersen et al., 1985; Rizzolatti et al., 1987] and visuospatial attention irrespective of eye-movements [Colby and Goldberg, 1999; Corbetta et al., 1998]. Recent electrophysiological works associated the posterior parietal cortex (PPC) with the processing of visual salience [Balan and Gottlieb, 2006; Constantinidis, 2006]; and suggested that neurons in the PPC represent the attentional spatial priorities irrespective of oculomotor behaviour [Bisley and Goldberg, 2003]. Moreover, the parietal cortex includes several areas with neurons that respond to stimuli in more than one modality [Avillac et al., 2007; Gross and Graziano, 1995], and that have been implicated in spatial orienting towards non-visual stimuli [Linden et al., 1999; Mazzoni et al., 1996; but see also Grunewald et al., 1999].

In this study, we investigated AV spatial interactions in the context of both overt orienting (eye-movements allowed) and covert orienting (central fixation). The overt orienting condition provided us with a naturalistic viewing situation, and enabled us to use gaze-position to investigate the impact of visual/auditory salience and AV spatial correspondence on orienting behaviour (i.e., an index of 'efficacy' that was then used for the fMRI analyses, cf. also Nardo et al., 2011). The inclusion of the covert viewing condition provided us with a data set where all subjects received exactly the same visual stimuli on the retina (which instead changed for each subject, as a function of gaze-position in the overt condition), and enabled us to generalise our fMRI findings to different (overt/covert) modes of spatial orienting.

Accordingly, during fMRI we presented subjects with short videos of naturalistic scenes. Each video contained a VE in the left or right visual hemifield, and an AE also presented either on the left- or on the right-hand side of space. This generated conditions with spatially 'congruent' or 'incongruent' AV stimuli, but always with full-field visual input and sounds matching with the objects/events in the visual environment (even when on the opposite side of the primary VE). In separate fMRI-runs, subjects viewed the stimuli either with central fixation required (covert orienting) or with eye-movements allowed (overt orienting), without receiving any further task-instruction.

The aims of the fMRI analyses were to highlight the blood oxygenation level-dependent (BOLD) correlates of stimulus-driven auditory salience, to assess any effect of AV spatial correspondence as a function of the strength (salience) of the visual and auditory signals and to investigate the relationship between these stimulus-driven factors and spatial overt/covert orienting. Our main hypothesis was that the saliency of visual and auditory signals would influence the spatial interaction between the two modalities, thus linking the processing of multisensory signals in complex dynamic environments with spatial aspects of selective attention. At the physiological level, the primary candidates to mediate any such effect were the parietal cortex and/or sensory areas in the occipital and superior temporal cortex (STC).

METHODS

Subjects

Twenty-six healthy right-handed volunteers (12 males, age range: 19–37 years; mean age: 26 ± 4.1) took part in the fMRI experiment. All participants were free of psychotropic or vasoactive medication, with no past history of psychiatric or neurological diseases. All had normal or corrected-to-normal (contact lenses) visual acuity, as well as self-reported normal hearing. After having received instructions, all participants gave their written informed consent. The study was approved by the independent Ethics Committee of the Santa Lucia Foundation (Scientific Institute for Research Hospitalization and Health Care).

Rationale and Design

This study was aimed at investigating the BOLD correlates of spatial orienting in ecologically valid AV conditions. We created a set of short videos showing everyday life situations and we manipulated four main variables: (i) the side of the primary VE (left, right hemifield); (ii) the presence of the AE (present, absent); (iii) the spatial correspondence between visual and AEs (same-side, opposite-side) and (iv) the mode of spatial orienting (overt, covert; i.e., with or without eye-movement allowed). Furthermore, we computed several indexes characterising the contribution of stimulus-driven signals (visual and auditory saliency maps, Vsal and Asal, for details, see the following sections) and the efficacy of AV events for spatial orienting (Eff, computed from subjects' gaze-positions, see the following sections). These indexes were then used as additional explanatory variables for the fMRI analyses.

Stimuli

Stimuli consisted of a set of 120 short videos (2.5 s each) displaying everyday situations in which one or more actors interacted with some objects within the environment (i.e., realistic context). Each stimulus contained a VE and an AE. The VE was associated with either the action of an agent (e.g., the actor puts an object on the table) or the setting on/off of a device (e.g., the TV was switched on). The AE was produced either by the action carried out by an actor (e.g., the noise of the object hitting the table) or emitted by an object present in the scene (e.g., computer, mobile phone, radio, TV, etc.).

The VE occurred on either the left-hand or right-hand side of the scene. The AE was also presented on the left or right (see below), either on the same side of the action/device producing the VE (spatially congruent AV conditions) or on the opposite side (incongruent AV conditions). Both VE and AE took place approximately 1 s after the video onset.

The crossing of the side of the VE and AE resulted in 'same-side' AV trials (CON, spatially congruent) and 'opposite-side' trials (INC, spatially incongruent). Moreover, in same-side trials, VE and AE could be either semantically/causally related (i.e., generated by the same agent/device) or unrelated (i.e., generated by the different agents/objects, but still on the same side). This further differentiation on spatially congruent trials dissociated the effect of AV spatial correspondence (common to 'related' and 'unrelated' conditions) from other semantic aspects that concerned the 'related' conditions only. Overall, the set of 120 stimuli included 80 congruent trials (40 of which AV related) and 40 incongruent trails. Within each category, 50% of the VEs included a human agent and 50% a device, balanced for left-/right-hand side of presentation.

Videos were shot in High Definition (1,920 × 1,080) with a digital camcorder (Canon Legria HF-S21) mounted onto a tripod. Sounds were recorded live-on-scene by means of a stereophonic microphone (Canon DM-100) mounted onto the camcorder, with a sampling rate of 44.100 Hz. Video editing was performed with Final Cut Pro 7.0 and sound editing with Soundtrack Pro 3.0, both running under Mac OS X Server 10.6. Videos were saved in .AVI uncompressed format with a resolution of 800 × 600 pixels and a frame rate of 25 Hz. The auditory stimuli were presented unilaterally either on the left- or right-hand side of external space (for details, see the next section).

Procedure

Subjects underwent two fMRI acquisition runs lasting about 12 min each. Each fMRI-run included the

presentation of all 120 videos. However, half of the stimuli were presented in 'sound' version ('S': including V and A events), whereas the other half included only the visual stimuli (no-sound version; 'NoS'). This enabled us to investigate the contribution of visual salience in the presence or absence of sounds (see below). The videos presented in the 'S' or 'NoS' conditions were counterbalanced across subjects, so that each of the 120 video was presented to 13 subjects with sound (S) and to the other 13 subjects without sound (NoS). For each subject, the order of presentation of the 120 stimuli was randomised within fMRI-run, but maintaining the same pool of S/NoS stimuli in the two runs.

In the first run ('covert session'), subjects were instructed to fixate the centre of the video, thus assessing activations associated with covert spatial orienting. To facilitate compliance with this instruction, a central fixation cross was presented at the centre of the visual display. In the second run ('overt session'), subjects were told that they could move their eyes freely and explore the scenes as they would do in a real environment. To minimise any between-subjects variability, the covert session was always presented first. Hence, only fMRI data pertaining to covert attention relate with the processing of the complex AV stimuli 'at first sight'.

Inside the magnetic resonance (MR) scanner, videos were back-projected onto a screen at the back of the MR bore that was visible to the subjects via an MR-compatible mirror. The screen covered a visual angle of approx. 20 × 15°. Sounds were delivered in external space in the proximity of the left-/right-hand side of the visual display. Two plastic tubes were positioned inside the MR head-coil, delivering sounds in a more realistic way (i.e., external locations) as compared with using headphones. Outside the MR-room, the tubes were connected to two loudspeakers and the volume was adjusted so that the sounds could be clearly heard against the MR-scanner noise. The sound delivery apparatus in the MR-coil was covered with black-cloth and was not visible to the subjects during the experiment.

Visual and Auditory Salience Indexes

We characterised the strength of stimulus-driven signals using computational models of visual and auditory salience. Briefly, for the visual modality, saliency maps were computed by using local centre-surround contrasts separately for intensity, colour, orientation, motion and flicker. This generates a series of conspicuity maps that were then combined into a unique saliency map by equally weighting each visual feature [Itti and Koch, 2001; Itti et al., 1998]. The resulting saliency map displays the most salient locations within a bi-dimensional space, representing the vertical and horizontal axes of a given visual stimulus (here, for each frame of the video). Asal maps were computed using an analogous approach, but now extracting centre-surround contrasts from a time–frequency spectrogram of the auditory signals of each video. Contrasts were computed separately for intensity, frequency, time and orientation [Kalinli and Narayanan, 2009; Kayser et al., 2005]. The resulting Asal map (equal weights for the four features) is also bi-dimensional, but now highlighting the occurrence of salient AEs in the time–frequency space. Our current implementation of the saliency models was modified from the software available at: http://www.saliencetoolbox.net.

Vsal and Asal maps were further processed to compute indexes of stimulus-driven attention that were then used as parametric modulators of event-related responses to VEs and AEs in the fMRI analyses. We considered a temporal window of 1.5 s, starting 1 s after the video onset (Fig. 1). The exclusion of the first second enabled us to minimise the contribution of the static frames that did not contain any VE; plus any centre-bias effect in the computation of the Eff index (cf. below; see also Tseng et al., 2009).

The Vsal index was computed as the ratio between the salience of the hemifield of the VE and the salience of the opposite hemifield. For each of the 120 stimuli, on a frame-by-frame basis, we extracted the mean salience separately for the two hemifields excluding a central area of 2° (see also section about Eff index, below). These values were averaged in the 1.5 s temporal window, and the difference was normalised to obtain an index between 1 and −1 (e.g., $[L - R]/[L + R]$, for a left VE). Positive values of Vsal indicate that the VE produced a stimulus-driven spatial bias towards the side of the VE; by contrast, negative values indicate that—despite the VE—salience was larger in the hemifield opposite to the VE. Accordingly, Vsal indexes to what extent the VE succeeded in producing a stimulus-driven bias on the side of the VE.

The Asal index was calculated by extracting the maximum value across frequencies for each time-point of the Asal map, and by averaging these values in the 1.5 s temporal window. For each of the 120 stimuli, Asal indexes the overall (i.e., non-spatial) stimulus-driven strength of the auditory input. As each sound was presented from a single external location (left- or right-hand side), we could associate the strength of the auditory input with one or the other hemifield. This enabled us to categorise bimodal AV trials as 'congruent' or 'incongruent', and to investigate the impact of auditory salience specifically on AV spatial interactions.

Eye-Movements Recording and Analysis (Eff index)

The horizontal and vertical gaze-position was recorded during both fMRI sessions with a long-range eye-tracking system compatible with the use in the MRI scanner (Applied Science Laboratories, Bedford, MA; Model 504; sampling rate = 60 Hz). Eye-tracking data recorded in the

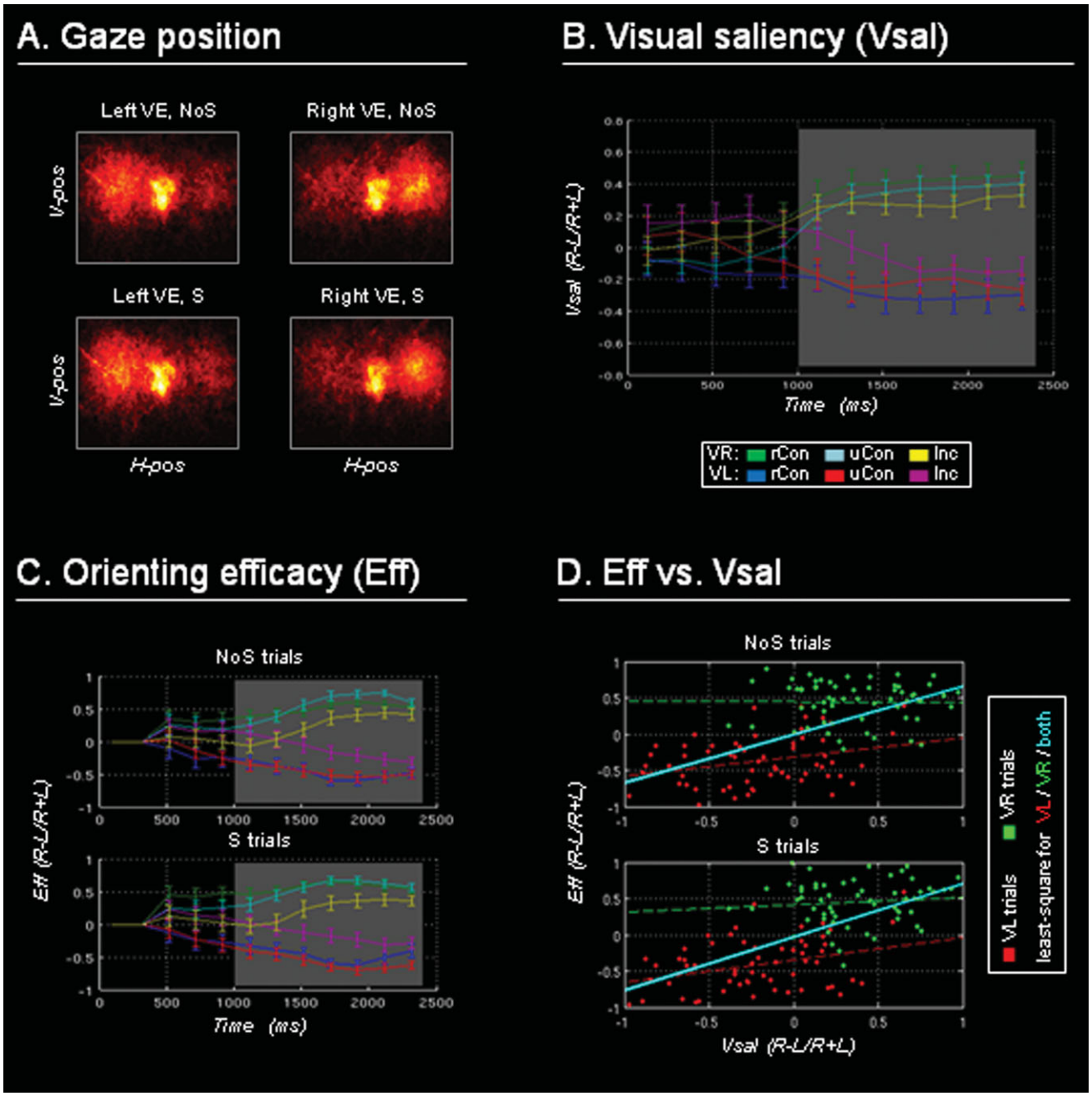

Figure 1.

Visual salience and gaze-data from the overt fMRI run. **A**: Gaze position for all subjects and all trials plotted separately for videos including left/right VEs, and presented without/with accompanying sounds (NoS/S). Subjects spatially oriented towards the side of the VE irrespective of sound presence. H-pos/V-pos: horizontal/vertical gaze position. **B**: Vsal bias (mean, ± s.e.m.; **METHODS** section) plotted over time and separately for the different video-types: VE on the left/right (VL/VR), with spatially congruent and related sounds (rCon); spatially congruent but unrelated sounds (uCon); and opposite-side, incongruent sounds (Inc). The VEs (approx., 1 s after video onset) led to a Vsal bias towards the corresponding side of the image. For the parametric analyses of fMRI data, we computed a Vsal index, averaging values between 1 and 2.5 s (see grey box). **C**: Gaze position data (ratio of time spent in the L/R hemifield; mean, ± s.e.m.) plotted over time for the different video-types, and separately for videos presented without/with sounds (NoS/S). Subjects shifted gaze towards the side of the VE, with analogous patterns irrespective of the presence of the sound. Average data in the 1–2.5 s window (grey box) were used as the Eff index (**METHODS** section). **D**: Vsal and Eff plotted against each other, for videos without/with sounds (NoS/S). The least-square regression lines highlight the relationship between these two measures (in cyan, considering all trials). This relationship was weaker when considering separately trials with right VEs (green) or left VEs (red), see corresponding (dotted) least-square lines. This suggests that factors other than bottom-up salience also contributed to spatial orienting behaviour (**RESULTS** section and **DISCUSSION** section).

covert fMRI-run served to check that subjects maintained central fixation and to identify unwanted saccades (modelled separately in the fMRI analyses, see below). Eye-tracking data recorded in the overt, free-viewing run were used to characterise the efficacy of VEs (and AEs, on bimodal trials) for triggering spatial orienting towards one or the other hemifield (Eff index).

The Eff index was computed separately for videos presented with (S) and without (NoS) sounds. This comprised three steps. First, for each subject, we extracted the time spent attending to the left and right hemifields of the display, during the 1.5 s temporal window (L_{time} and R_{time}). For this, we excluded any data-point falling into a 2° central area, because small deviations of horizontal gaze-position around the centre of the screen (even below the spatial precision of our measurement) would inappropriately affect the L_{time}/R_{time} values (Fig. 1A). Next, we computed the difference between the time spent in the two hemifields and normalised this between 1 and −1 (e.g., $[L_{time} - R_{time}]/[L_{time} + R_{time}]$, for a left VE). Finally, for each of the 120 stimuli, we averaged individual ratios across subjects, separately for the 13 subjects who were presented with the sound versions (S) and the 13 subjects presented with the silent versions (NoS) of the videos. Accordingly, for each of the 120 videos we generated two Eff indexes corresponding to the tendency of the subjects to look towards the VE, in the presence or the absence of the corresponding AE.

fMRI Acquisition

A Siemens Allegra (Siemens Medical Systems, Erlangen, Germany) 3T scanner equipped for echo-planar imaging (EPI) was used to acquire functional MR images. A quadrature volume head coil was used for radio frequency transmission and reception. Head movement was minimised by mild restraint and cushioning. Thirty-two slices of functional MR images were acquired using BOLD imaging (3 × 3 mm, 2.5 mm thick, 50% distance factor, repetition time = 2.08 s, time echo = 30 ms), covering the entirety of the cortex.

fMRI Analysis

Data pre-processing and analysis were performed with SPM8 (Wellcome Department of Cognitive Neurology). A total of 710 fMRI volumes for each subject were acquired in two separate runs (2 × 355). After having discarded the first four volumes of each run to account for T1 saturation effects, images were realigned in order to correct for head movements and slice-timed to the middle slice as reference. Images were normalised to the MNI EPI template, re-sampled to 2-mm isotropic voxel size and spatially smoothed using an isotropic Gaussian kernel of 8-mm full width at half maximum.

All fMRI analyses included first-level within-subject analyses and second-level analyses for random effects statistical inference at the group level [Penny and Holmes, 2004]. The aim of the fMRI analysis of the first run was to investigate the correlates of AV spatial processing in covert attention, whereas in the second run we aimed at investigating these effects in overt spatial orienting mode.

For both fMRI-runs, the stimuli were modelled as delta functions time-locked 1.5 s after the video onset (duration = 0), convolved with the standard SPM hemodynamic response function. The subject-specific models included 12 event-types given by the crossing of the factors: side of the VE (left vs. right), presence of sound (AE present vs. absent), and AV spatial correspondence (same-side vs. opposite-side, the former further divided into related and unrelated VE/AE).

To assess the impact of visual salience, auditory salience and orienting efficacy, each model included parametric regressors corresponding to Vsal, Asal and Eff indexes, separately for each of the six event-types with sound present (S trials). The Vsal and Eff indexes were also included as modulators of the six event-types without any sound (NoS trials). For the covert fMRI-run, the first-level models took into account any loss of fixation (except for two subjects with poor quality eye-tracking data). Trials containing any change of eye-position larger than 1.5° and lasting longer than 100 ms were modelled as a separate event-type and discarded from subsequent group-level analyses (2.6% of trials across subjects). All models included the head-motion realignment parameters as additional covariates of no interest. The time-series were high-pass filtered at 128 s and pre-whitened by means of autoregressive model AR(1).

Group-level analyses were carried out separately for the covert and overt sessions, but using analogous second-level statistical models. Repeated-measures ANOVAs were used to analyse the event-related activations associated with the 12 main event-types (see above), whereas the effects of the parametric regressors were tested using separate one-way ANOVAs. ANOVAs concerning Vsal and Eff indexes included 12 conditions corresponding to the modulatory effects of visual salience and Eff for each of the 12 event-types. ANOVAs concerning Asal included only six conditions corresponding to the event-types with auditory stimulation. Sphericity-correction was applied in all models to account for possible differences in error variance across conditions and any nonindependent error term for repeated measures [Friston et al., 2002].

For all analyses, we report activations corrected for multiple comparisons at the cluster level, considering the whole brain as the volume of interest (FWE P-corr. <0.05; cluster size estimated at a voxel-level threshold of P-unc. $= 0.001$). The localization of the activation clusters was based on the anatomical atlas of the human brain by Duvernoy (1991).

RESULTS

Behavioural Data

In the overt fMRI run, subjects' gaze-direction was used as a measure of the deployment of spatial attention.

Overall, the subjects showed a tendency to orient towards the side of the VEs, irrespective of the presence or absence of the sound (Fig. 1A). Figure 1B shows that the presentation of the VE caused a bias of visual salience towards the side of the VE; and Figure 1C shows the corresponding bias in the time that subjects spent looking on the side of the VE (both plots displaying these effects over time, and as a function of the AV condition). The Vsal bias was found to reliably predict the gaze-direction bias (Eff), irrespective of the presence or the absence of sounds (top and bottom plots, Fig. 1D). Specifically, we found significant linear relationships between Vsal and Eff both in the absence of sounds (NoS conditions, $r = 0.58$, $P < 0.001$; top in Fig. 1D, least-square line in cyan) and in the presence of sounds (S condition, $r = 0.61$, $P < 0.001$; bottom in Fig. 1D, in cyan). This relationship was weaker when considering separately trials in the two hemifields, suggesting that factors other than visual salience contributed to spatial orienting (**DISCUSSION** section). The correlation between saliency and efficacy was still significant for VEs on the left-hand side (NoS: $r = 0.30$, $P < 0.022$; S: $r = 0.33$, $P < 0.010$; see Fig. 1D in red), but not for VEs on the right-hand side (NoS: $r = -0.01$, $P > 0.932$; S: $r = 0.10$, $P > 0.430$; see Fig. 1D in green).

Imaging Data

We first report the results related to the event-related responses associated with the presentation of the VEs and AEs, and then turn to the trial-specific parametric effects related to Asal, Vsal and Eff. Overall, the results were highly consistent in covert and overt fMRI-runs (data are presented side-by-side in Figs. 2–4 and Tables I and II). We detail any difference between the two modes of orienting in the last paragraph of the **RESULTS** section.

Event-Related Responses to VEs and AEs

We examined brain activations associated with the presentation of complex sounds (AEs), the effect of side of the VEs and the spatial correspondence between events in the two modalities. It is important to note that in the context of this study, we refer to 'left' (L) and 'right' (R) visual conditions depending on the side of the VE, but all stimuli/videos comprised visual stimulation on both sides.

The complex sounds ('S' – 'NoS' trials) activated the STC, including Heschl's gyri (i.e., primary auditory cortex), with the clusters extending into the superior temporal sulcus and the middle-temporal gyrus (Fig. 2A and Table I). The activation of this region was bilateral irrespective of L/R sound position and the position of the VE (see signal plots in Fig. 2A).

The effect of the side of the VEs ('VEleft' – 'VEright' and 'VEright' – 'VEleft') revealed activation in the occipital and posterior parietal cortex (PPC) contralateral to the side of the VE. In the occipital cortex, clusters of activation included the superior occipital gyrus, the inferior occipital gyrus and the lateral occipital gyrus plus the adjacent middle-temporal complex (MT+). The involvement of MT+ was confirmed using the 'Anatomy Toolbox 1.8' for SPM. This showed that the whole of MT+ was activated bilaterally in the covert fMRI-run, and that 89% (right hemisphere) and 43% (left hemisphere) was activated in the overt run. In the parietal lobe, we found activation in the PPC, extending into the medial wall of the intraparietal sulcus (IPS; Fig. 2B and Table I). Accordingly, although the videos contained complex visual stimuli on both sides, the event-related analysis successfully detected activations associated with the lateralised VE.

Next, we investigated the effect of AV spatial correspondence by comparing conditions with 'congruent' (same-side) minus 'incongruent' (opposite-side) AV events. Somewhat surprisingly, this did not reveal any significant modulation of the event-related responses. Nonetheless, the effect of AV spatial correspondence became apparent as soon as we considered the level of salience of the AE (see the next section).

Modulations Related to Auditory Salience (Asal)

This analysis tested whether the activation associated with the AE was modulated depending on the salience of the auditory input (Asal), and whether any such effect changed as a function of the AV spatial correspondence. Irrespective of AV spatial correspondence, we found that activity in the auditory cortex increased with increasing values of auditory salience (Fig. 3A and Table II). The salience-modulated region was found within the region showing an overall event-related response to the AE (Figs. 3A and 2A), and included primarily the superior temporal gyrus posterior to the Heschl's gyrus (i.e., planum temporale). Note that this modulatory effect was found after accounting for the overall (mean) response to the AE (event-related response in the previous section). The signal plots in Figure 3A show the parameter estimates associated with the modulatory effect of Asal. These correspond to the slopes of the linear relationship between the auditory salience and the amplitude of the event-related BOLD response to the sounds, separately for left-/right-hand side AEs and spatially congruent/incongruent AV conditions. The slopes were more positive for left-hand side sounds than right-hand side sounds, but this difference was not statistically significant.

Next, we tested whether the spatial relationship between the events in the two modalities affected brain activations associated with auditory salience on AV trials (i.e., congruent vs. incongruent). This revealed that activity in the PPC increased with increasing auditory salience when the videos included same-side AV events, but decreased with increasing auditory salience when visual and AEs were on opposite sides (Fig. 3B). The effect was significant after

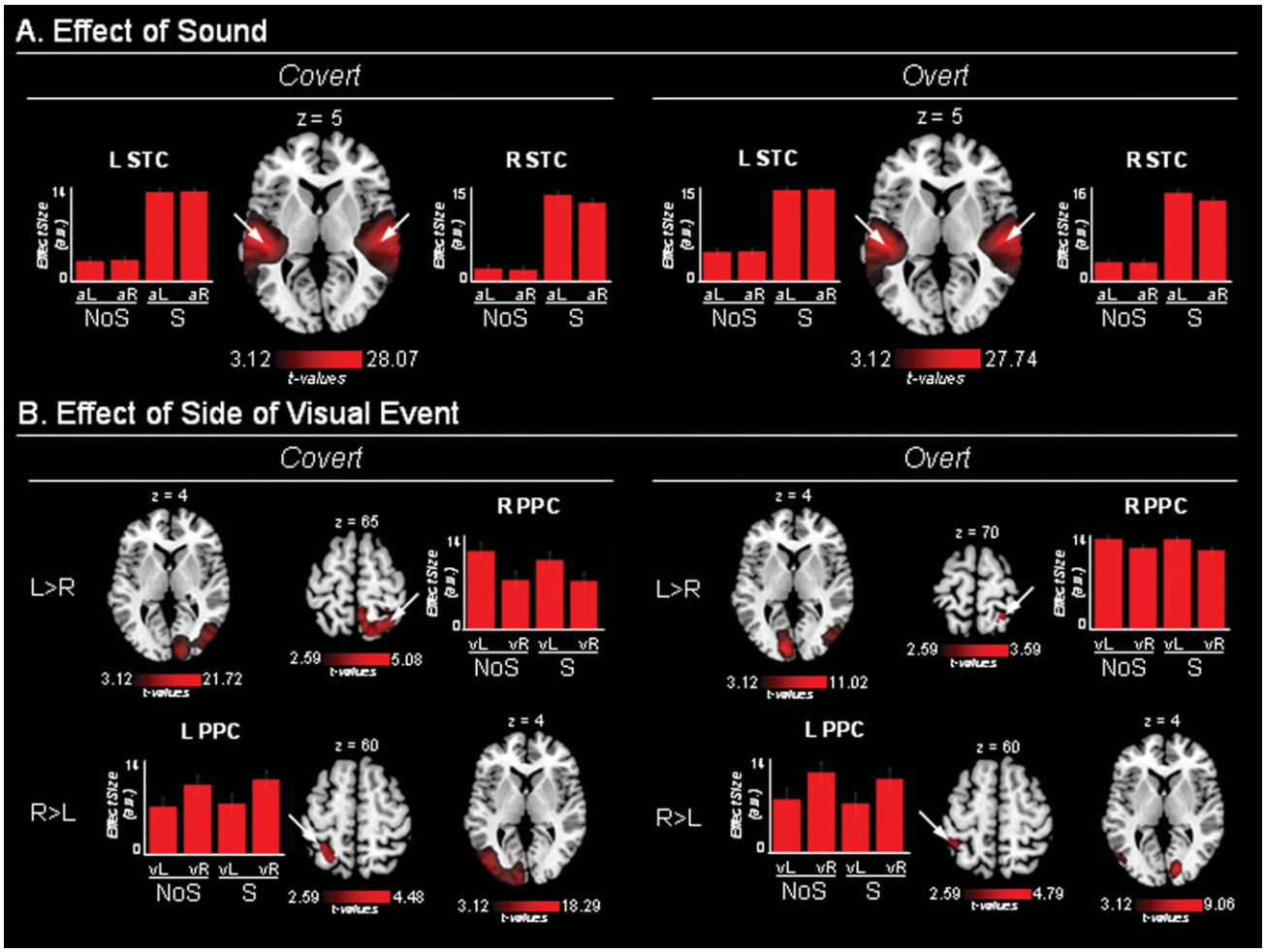

Figure 2.

Event-related responses to naturalistic stimuli. **A**: Brain regions showing event-related activation time-locked to complex AEs. This revealed significant activation in bilateral STC. aL/aR: left-/right-hand side of the AE, irrespective of AV spatial alignment that was not significant in any of these areas. **B**: Brain regions showing event-related activation time-locked to complex/dynamic VEs. In the covert orienting condition (left panels), this revealed significant activation in striate and extrastriate visual cortex (including MT+), and in the PPC contralateral to the side of the VE. In the overt condition (right panels), the activation of striate cortex was now ipsilateral to the VE, whereas remained contralateral in higher-order visual areas and PPC. vL/vR: left-/right-hand side of the VE, irrespective of AV spatial alignment that was not significant in any of these areas. NoS/S: videos presented without/with corresponding sounds. Activations are projected on the standard MNI template. Effect sizes are plotted in arbitrary units (a.u.); error bars are 90% confidence intervals.

correction for multiple comparisons at the whole-brain level only in the left hemisphere, for the covert session. Nonetheless, when we considered this cluster as a restricted volume of interest for the overt session (small volume correction; Worsley et al., 1996), and a symmetrical volume in the right hemisphere (overt and covert sessions), we found statistically significant effects in both hemispheres (overt session only, Table II). The clusters of activation extended from the superior parietal gyrus into the fundus of the IPS (descending segment). In the superior parietal gyrus, this effect partially overlapped the cluster, showing an event-related response for the 'side of the VE' (Figs. 3B and 2B).

We further examined the interaction between the AV spatial correspondence and the strength/salience of the AE by comparing AV-related and AV-unrelated conditions, but this contrast showed no significant result. This suggests that the effect of AV spatial correspondence in parietal cortex reflected processes related to multisensory space representation and/or supramodal attentional orienting, rather than the integration of object information across modalities.

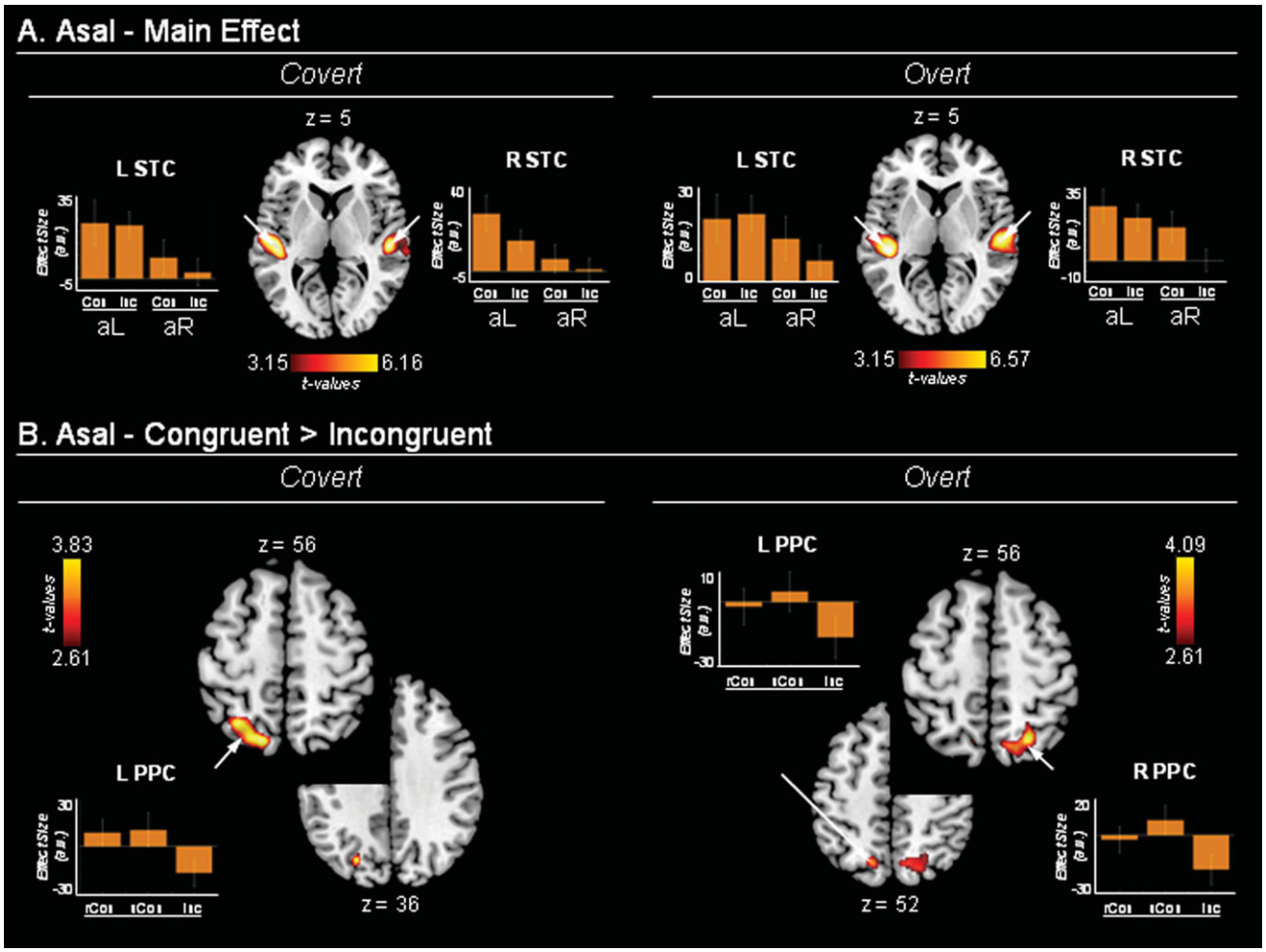

Figure 3.

Modulatory effects related to Asal. **A**: Brain regions where the BOLD signal co-varied positively with the Asal index. Significant effects were found in the STC bilaterally. The signal plots show the parameter estimates associated with the effect of Asal, separately for left and right sounds (aL/aR) and congruent/incongruent AV conditions. The plots show a larger effect of Asal for left than right sounds, but this was not statistically significant. **B:** In the PPC, we found an interaction between Asal and AV spatial alignment. The BOLD signal increased with increasing auditory salience when visual and AEs occurred on the same side of space (i.e., congruent trials, positive parameter estimates in the signal plots), but decreased when visual and AEs occurred on opposite sides (incongruent trials, negative parameter estimates). rCon/uCon: spatially congruent related/unrelated AV events; Inc: spatially incongruent AV events. Activations are projected on the standard MNI template. Effect sizes are plotted in arbitrary units (a.u.); error bars are 90% confidence intervals.

Finally, we further probed the relevance of the Asal index by re-analysing the fMRI data, this time using sound amplitude instead of salience. For each stimulus, we extracted the average sound power (Cool Edit Pro 2.0, Syntrillium, USA) and used this as a new trial-by-trial parametric regressor. The group analyses showed a significant amplitude effect in the superior temporal gyrus bilaterally ($x,y,z = 56,-16,4$, $t = 7.27$, P-corr. <0.001; and, $x,y,z = -52,-16,-2$, $t = 6.48$, P-corr. <0.001), irrespective of AV congruence and side. By contrast, the new amplitude-related regressor did not show any crossmodal spatial effect in the PPC (i.e., no interaction between amplitude and AV spatial alignment), even when considering only the PPC as the volume of interest (maximum peak at: $x,y,z = -28,-58,54$, P-unc. >0.05, P-SVC-corr >0.689). These additional analyses indicate that sound amplitude can partially explain the effect of salience in the auditory cortex (see also **DISCUSSION** section), but cannot account for the crossmodal spatial effect related to auditory salience in the PPC.

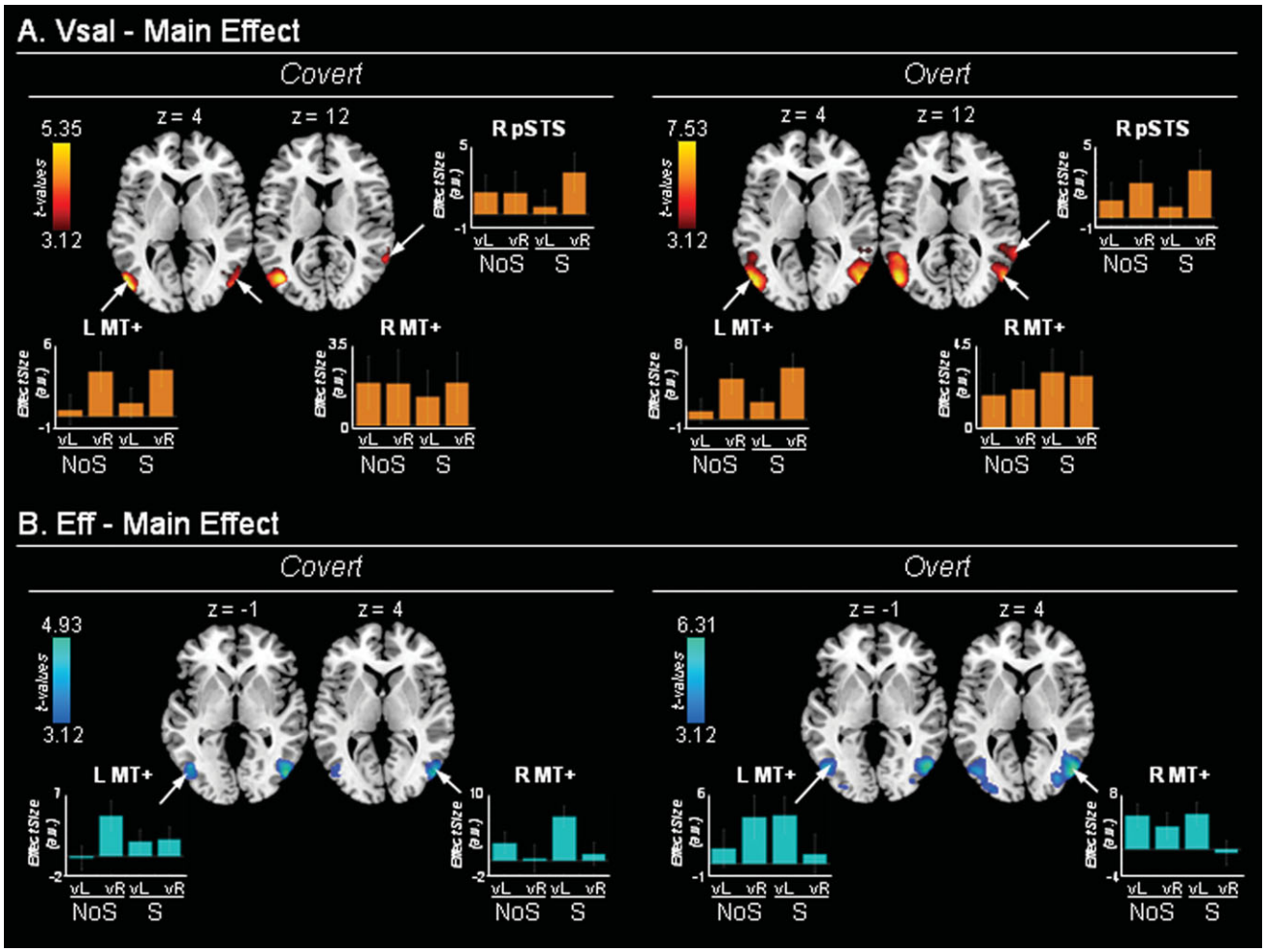

Figure 4.

Modulatory effects related to Vsal and Eff. **A**: Brain regions where the BOLD signal co-varied positively with the Vsal index. Significant modulations were found in the middle-temporal complex (MT+) bilaterally, and in the right posterior superior temporal sulcus (pSTS). In the left MT+, this modulatory influence also showed a significant effect of side, with a larger modulation when the VE was presented on the (contralateral) right-hand side as compared to the (ipsilateral) left-hand side, see parameter estimates in the corresponding signal plots and Table II. **B**: Brain regions where the BOLD signal co-varied positively with the Eff index. Significant modulations were found in the middle-temporal complex bilaterally (MT+), with a substantial overlap between the effects of Vsal and Eff (A and B). This indicates that activity in MT+ reflected both stimulus-driven signals and their efficacy for spatial orienting. vL/vR: left-/right-hand side of the VE, irrespective of AV spatial alignment that was not significant in any of these areas. NoS/S: video presented without/with corresponding sound. Activations are projected on the standard MNI template. Effect sizes are plotted in arbitrary units (a.u.); error bars are 90% confidence intervals.

Modulations Related to Visual Salience (Vsal)

Concerning visual salience, we first tested for areas where activity increased with increasing salience irrespective of the side of the VE, and then looked for effects that were specific depending on the side of the VE. The overall effect of Vsal revealed modulation in bilateral MT+ complex, and in the right posterior superior temporal sulcus (pSTS; Fig. 4A and Table II). This modulatory effect of Vsal in MT+ partially overlapped with the most anterior part of the activation found in the event-related analyses (main effect of 'side of the VE', cf. Fig. 2A).

In the left MT+, salience was found to modulate visual responses only during the presentation of right-hand side VEs; by contrast, in the right MT+ (and right pSTS), salience modulated VE responses irrespective of side (signal plots, Fig. 4A; Table II). No further modulation by the presence/absence of sound or the congruence/incongruence of the AV stimuli was found.

TABLE I. Event-related responses to auditory and visual naturalistic events

		Covert					Overt				
		Cluster			Peak		Cluster			Peak	
Contrast	Region	*P*-cor	*k*	[%]	*t*-Value	*x,y,z*	*P*-cor	*k*	[%]	*t*-Value	*x,y,z*
S > NoS	L superior temporal cortex	<0.001	8,112		28.07	−48,−26,6	<0.001	7,629		27.74	−52,−24,4
	R superior temporal cortex	<0.001	8,448		26.48	56,−18,2	<0.001	7,681		26.57	56,−18,2
L > R	R ventral occipital cortex	<0.001	13,459		21.72	16,−82,−12	<0.001	5,253		5.15	26,−78,−14
	R dorsal occipital cortex				14.02	24,−92,26				5.86	28,−84,34
	R middle-temporal complex (MT+)			[100]	11.57	46,−76,4			[89]	6.72	46,−72,10
	R posterior parietal cortex (PPC/IPS)				5.08	20,−60,70	*n.s.*			*3.59*	*24,−46,70*
	L calcarine cortex	—	—		—	—	<0.001	2,646		11.02	−10,88,4
R > L	L ventral occipital cortex	<0.001	8,878		18.29	−16,82,−14	0.013	318		3.70	−48,−72,−8
	L dorsal occipital cortex				14.68	−20,−94,22	*n.s.*			*3.42*	*−16,−86,42*
	L middle-temporal complex (MT+)			[100]	11.64	−42,−82,8			[43]	4.69	−52,−70,4
	L posterior parietal cortex (PPC/IPS)	0.005	389		4.04	−26,−56,60	*0.094*	184		4.79	−40,−46,66
	R calcarine cortex	—	—		—	—	<0.001	680		9.06	14,−84,6

S > NoS: brain regions activated by videos presented with sounds, as compared to videos presented without sounds. L > R: brain regions activated by videos containing VEs on the left-hand side, as compared to videos containing VEs on the right-hand side. R > L: brain regions activated by videos containing VEs on the right-hand side, as compared to videos containing VEs on the left-hand side. *P*-values are corrected for multiple comparisons at the whole-brain level (except for *t*-values reported in *italics*), and *k* is the number of voxels in each cluster. [%]: percentage of MT+ activated by the cluster (as determined by the 'Anatomy Toolbox 1.8'). Overt/Covert: spatial orienting with/without eye movements. IPS: intraparietal sulcus.

TABLE II. Modulations related to auditory salience, visual salience and orienting efficacy

		Covert					Overt				
		Cluster			Peak		Cluster			Peak	
Contrast	Region	*P*-cor	*k*	[%]	*t*-Value	*x,y,z*	*P*-cor	*k*	[%]	*t*-Value	*x,y,z*
Asal											
ME	L superior temporal cortex	<0.001	1,196		6.58	−54,−18,2	<0.001	1,422		7.09	−44,−30,8
	R superior temporal cortex	<0.001	903		6.14	52,−22,4	<0.001	1,480		6.55	54,−18,2
CON > INC	L posterior parietal cortex (PPC/IPS)	0.038	191		3.76	−16,70,58	0.030*			3.19	−14,−68,52
	R posterior parietal cortex (PPC/IPS)	—	—		—	—	0.003*			3.91	24,−62,58
Vsal											
ME	L middle-temporal complex (MT+)	0.001	573	[55]	5.35	−52,78,6	<0.001	1,531	[75]	7.53	−52,−72,8
	R middle-temporal complex (MT+)	0.006	429	[59]	4.21	48,−76,2	<0.001	1,444	[99]	6.28	52,−74,6
	R posterior superior temporal sulcus (pSTS)				4.00	60,−50,14				4.87	64,−48,12
	L posterior parietal cortex (PPC/IPS)	*n.s.*			*2.28*	*−34,−40,56*	—		—	—	—
	R posterior parietal cortex (PPC/IPS)	—	—		—	—	*n.s.*			*2.04*	*24,−44,64*
L > R	L/R calcarine cortex	—	—		—	—	<0.001	3,176		5.59	−10,96,−2
R > L	L middle-temporal complex (MT+)	0.011	375	[13]	4.50	−44,−74,14	*n.s.*		*[0.2]*	*4.29*	*−52,−74,10*
Eff											
ME	L middle-temporal complex (MT+)	0.016	302	[60]	4.39	−44,−68,0	<0.001	2,250	[91]	5.33	−50,−72,0
	R middle-temporal complex (MT+)	0.001	553	[92]	4.93	50,−74,0	<0.001	2,346	[97]	6.31	50,−70,2
L > R	R ventral occipital cortex	<0.001	2,569		8.29	16,−80,−10	—		—	—	—
	R calcarine cortex				5.46	12,−90,2	*n.s.*			*4.04*	*−26,−52,−10*
	R middle-temporal complex (MT+)	0.003	423	[90]	5.73	46,−70,8	*n.s.*		*[21]*	*3.55*	*48,−72,0*
	L calcarine cortex	—	—		—	—	<0.001	705		7.01	−10,−88,6
R > L	L ventral occipital cortex	<0.001	2,582		7.25	−34,−80,−16	*n.s.*			*3.33*	*−36,−88,−8*
	L middle–temporal complex (MT+)			[77]	6.61	−48,−80,0	—		—	—	—
	R calcarine cortex	—	—		—	—	<0.001	1,017		7.51	12,−86,4

Asal, main effect (ME): brain regions where BOLD signal co-varied positively with auditory salience. Asal, CON > INC: regions where BOLD signal co-varied positively with auditory salience only when V and A events were spatially aligned (i.e., occurred on the same side of space). Vsal, ME: brain regions where BOLD signal co-varied positively with visual salience. Vsal, L > R: regions where BOLD signal co-varied positively with visual saliency for videos with VEs on the left, as compared to videos with VEs on the right. Vsal, R > L: regions where BOLD signal co-varied positively with visual saliency for videos with VEs on the right, as compared to videos with VEs on the left. Eff, ME: brain regions where BOLD signal co-varied positively with the efficacy of spatial orienting. Eff, L > R: regions where BOLD signal co-varied positively with visual orienting efficacy for videos with VEs on the left, as compared to videos with VEs on the right. Eff, R > L: regions where BOLD signal co-varied positively with orienting efficacy for videos with VEs on the right, as compared to videos with VEs on the left. *P*-values are corrected for multiple comparisons at the whole-brain level (except for *t*-values reported in *italics*), and *k* is the number of voxels in each cluster. [%]: percentage of MT+ activated by the cluster (as determined by the 'Anatomy Toolbox 1.8').

P*-values are small-volume-corrected within a parietal Region-of-Interest (METHODS** section). Overt/Covert: spatial orienting with/without eye movements. IPS: intraparietal sulcus.

Modulations Related to the Efficacy of Spatial Orienting (Eff)

Finally, we considered the modulatory effect associated with our Eff index. The Eff index was calculated on the basis of eye-tracking data recorded during the overt session, and then used for the fMRI analyses of both overt and covert sessions. Irrespective of the side of the VE, we found that activity in MT+ increased with increasing Eff; that is, the longer the time subjects spent looking towards the side of the VE, the greater the activation in MT+ (Fig. 4B and Table II). The direct comparison between left-hand side and right hand side VEs showed that the modulatory effect of Eff tended to be larger in the hemisphere contralateral to the side of the VE (signal plots in Fig. 4B, statistically significant effects of side in the covert session only; Table II).

In MT+, the effect of Eff overlapped with both the event-related responses and the effect of visual salience (Figs. 2B, 4A and 4B). Accordingly, MT+ responded to lateralised VEs, did more so when the VE was salient, and activity there was further enhanced when the VE effectively attracted subjects' gaze/attention.

Results in Overt Condition

Overall, the results of analyses in the overt session (i.e., the second fMRI-run, with eye-movements allowed) substantially replicated the results found in covert condition (Tables I and II). One exception concerned the event-related effect of the 'side of the VEs'. In covert condition, this showed clusters of activation contralateral to the side of the VE in the occipital striate and extrastriate cortex, plus the PPC (Fig. 2B, panels on the left). By contrast, in overt condition, the activation of the striate cortex was ipsilateral to the side of the VE (Fig. 2B, panels on the right; Table I), whereas activity in extrastriate and parietal cortex remained contralateral to the VE.

DISCUSSION

We investigated brain responses associated with the processing of naturalistic AV stimuli in complex environments, where multiple sources of sensory information interact and compete to guide spatial orienting behaviour. We identified patterns of brain activity associated with the bottom-up strength of the VEs and AEs (indexed using computational models), and activations related to the efficacy of the VEs for spatial orienting (indexed using eye-movements). The processing of complex VEs was associated with the activation of the visual cortex and the PPC contralateral to the side of the event. The salience of the VE and its efficacy for spatial orienting further modulated activity in the occipito-temporal cortex. The salience of naturalistic sounds boosted auditory responses in bilateral auditory cortex, irrespective of sound side and the spatial correspondence between auditory and VEs. In the PPC activity increased with increasing auditory salience, but only when the sound was on the same side as the VE. This crossmodal spatial effect was found irrespective of whether subjects were allowed to move their eyes or not. Our results demonstrate that computational models of stimulus-driven attention can predict activity in sensory cortices during viewing of complex multisensory stimuli, and that the PPC combines spatial information about vision and audition in naturalistic conditions entailing high levels of sensory competition.

Overt Orienting in Complex AV Environments

Behaviourally, we found that visual salience predicted spatial orienting behaviour: the greater the salience bias induced by the VE, the longer the time subjects spent attending on that side (Fig. 1D). However, this effect was much smaller when we considered the effect of visual salience on gaze separately for left-hand side and right-hand side VEs. This indicates that the impact of salience on gaze was primarily due to the lateralised presentation of the VE, suggesting that the VE triggered processes other than pure bottom-up attention capture. These, in turn, led to the tendency of the subjects to attend for longer times on the side of the VE. This observation is consistent with the proposal that endogenous, object-related factors contribute to spatial orienting behaviour during exploration of complex visual scenes [Einhäuser et al., 2008; Nuthmann and Henderson, 2010].

A second aspect emerging from the analysis of gaze-data was that sounds, and AV spatial congruence, had no significant effect on overt orienting behaviour (Fig. 1C). This was somewhat surprising, because a previous behavioural study reported that sounds can influence gaze-direction during viewing of complex visual scenes [Onat et al., 2007]. However, there are several important differences between this previous study and our work. First, Onat and colleagues used visual stimuli that were not conceived to produce a systematic spatial bias towards one side, whereas this was a fundamental aspect of our videos (i.e., the VEs, see also Fig. 1A). Second, in Onat's study participants were asked to 'watch and *listen carefully* to images and *sounds*'. This specific task instruction, together with the presentation of auditory-only trials, may have led subjects to pay more attention to the sounds than in this study. Third, Onat's study used static pictures presented for a relatively long time (6 s), whereas here the subjects were presented with brief and dynamic visual stimuli (2.5 s, with the VEs lasting only 1–1.5 s). We believe that these differences biased processing in favour of the visual input—and the VEs in particular—in this study, leading to weak/no effect of sounds on spatial orienting behaviour (but see imaging results below).

Processing of VEs in the Occipital Cortex

Although the naturalistic videos contained complex visual stimuli in both hemifields, the event-related fMRI

analysis successfully identified activations associated with lateralised VEs. In the covert viewing condition, increased BOLD activity was found in the striate and extrastriate occipital cortex with larger activation for contralateral versus ipsilateral VEs. The clusters of activation in the lateral occipital cortex extended into the middle-temporal complex (MT+, a region known to be involved in the processing of visual motion/flicker, see Schiller, 1993), compatibly with the sensory characteristics of the VEs in this study (**METHODS** section). In the overt viewing condition, when subjects typically shifted gaze towards the VEs (see above, and Fig. 1C), activation of the striate cortex switched to the ipsilateral hemifield. This indicates that, in overt condition, the activation of striate cortex largely reflected the augmented visual input provided by the part of the image opposite to the VE (e.g., following gaze-shifts towards left VEs, most of the image fell in the right visual field, in retinal coordinates). By contrast, the pattern of activation in MT+ was unaffected by the overt/covert mode of orienting: activity increased in the hemisphere contralateral to the VE, even when subjects shifted their gaze towards the VE. This indicates that the side-specific BOLD responses in MT+ took into account the change in gaze-position [Bremmer et al., 1997; Leopold and Logothetis, 1998]. The differential impact of the orienting mode (overt/cover) on the activity of striate and extrastriate areas is consistent with hierarchical representations of the visual field that change progressively from 'retinotopic' in early visual areas to contralateral (with little or no topographic organization) in higher-order extra-occipital areas [Jack et al., 2007].

The large variety of the naturalistic visual stimuli enabled us to ask whether the stimulus-driven strength of the VE and the efficacy of the VE for spatial orienting further modulated these event-related responses, on a trial-by-trial basis. We used a computational model of visual attention [Itti and Koch, 2001; Itti et al., 1998] to quantify the strength of the stimulus-driven spatial bias produced by each VE. For each VE, we indexed the amount of stimulus-driven spatial bias by computing the difference between mean salience in the hemifield of the VE and mean salience in opposite hemifield (Fig. 1B). We found that the amount of spatial bias associated with the VE modulated activity in bilateral MT+ and in the right posterior superior temporal sulcus (pSTS; Fig. 4A). In the right hemisphere (right MT+ and right pSTS), the BOLD signal increased with increasing saliency bias irrespective of the side of the VE, whereas in the left hemisphere the salience of the VE affected MT+ primarily when the VE was presented on the (contralateral) left side. The modulatory effect of visual salience in MT+ was unaffected by the overt/covert orienting mode, suggesting that any side-specificity (here, left hemisphere only) was not strictly dependent on the VEs position in retinal coordinates.

In MT+ the event-related responses associated with the VE were further modulated by the efficacy of the VE in attracting subjects' gaze (Fig. 4B). This measure of the 'attention grabbiness' of the VE was found to modulate MT+ activity also in covert viewing conditions, consistent with an attentional—rather than oculomotor—origin of these effects. It should be stressed that our fMRI analyses included both salience and efficacy (Vsal and Eff) within the same multiple regression model, and that these indexes were computed separately for left and right VEs. This is important because a left VE tended to cause both leftward saliency bias and a left gaze-orienting bias, and vice versa for right VEs (Fig. 1D, and discussion of the behavioural data above). The results of our imaging analyses demonstrate that activity in MT+ was selectively modulated by both visual salience and orienting efficiency, after accounting for any effect common to the two factors.

These results support the view that stimulus-driven and endogenous factors jointly contribute to processing in the visual cortex. On the one hand, low-level aspects of the sensory input have been found to influence selection and competitive processes, both at the single-cell level [Buschman and Miller, 2007; Constantinidis and Steinmetz, 2005], and using neuroimaging techniques [Beck and Kastner, 2007; McMains and Kastner, 2010]. On the other hand, several lines of evidence demonstrated that top-down attention modulates the processing of incoming visual signals and the associated activity in the visual cortex [Corbetta et al., 1990; Kastner et al., 1999]. The 'biased competition model' [Desimone and Duncan, 1995] provides us with a framework to interpret the interplay between stimulus-driven and top-down factors during visual selection. Accordingly, the strength of the bottom-up signals determines which neuron populations activate, whereas the role of attention is to bias competition towards one of these populations. Both electrophysiological and neuroimaging studies provide support to this view [Kastner et al., 1999; Recanzone and Wurtz, 2000; Reddy et al., 2009]. Our results here show that the processing of naturalistic visual stimuli, including multiple competing signals, entails an interplay between bottom-up and top-down attentional mechanisms. This interplay affects processing in visual cortex (here, primarily MT+) boosting the BOLD response to VEs located in the 'most interesting' side of space.

Visual Responses and AV Spatial Interactions in the PPC

Together with these effects in the visual occipital cortex, we also found that the PPC participated to the processing of the naturalistic stimuli. The event-related analysis showed activation of the PPC with larger responses for contralateral than ipsilateral VEs. The clusters of activation extended from the superior parietal gyrus to the adjacent medial wall of the IPS (the putative human homologue of monkey's area LIP, see Grefkes and Fink, 2005; Koyama et al., 2004). These effects were observed irrespective of the overt/covert mode of orienting, in agreement with the view that the PPC/IPS contains spatial representations

encoded in a non-retinotopic frame of reference [Brotchie et al., 1995; Duhamel et al., 1997; Snyder et al., 1998], and consistently with the previous imaging findings, showing activation of the same frontoparietal network for covert and overt spatial orienting [Corbetta et al., 1998; Nobre et al., 2000]. Here, we did not find any significant modulation of activity in PPC/IPS by visual salience [cf. Bogler et al., 2011; Nardo et al., 2011] although a few clusters were found when lowering the statistical threshold (Table II, in *italics*).

By contrast, here we found that activity in the PPC was modulated by auditory salience and that mere trial-by-trial variability in sound amplitude could not explain this effect. Most importantly, the effect of auditory salience was strictly dependent on the spatial relationship between the VEs and the AEs. When the sound was presented on the same side of the VE, activity in the parietal cortex increased with increasing sound salience. By contrast, when the sound was on the opposite side of the VE, the activity in PPC decreased with increasing auditory salience (signal plots in Fig. 3B). This crossmodal spatial effect was found irrespective of overt/covert orienting mode (albeit only using a restricted volume of interest in the overt session), and did not depend on whether the sound was or was not semantically/causally related to the VE. This further emphasises the spatial nature of this crossmodal interaction. Anatomically, the effect involved the superior parietal gyrus, with the clusters extending ventrally into the fundus of the posterior IPS, possibly including the putative human homologue of monkey's area VIP [Bremmer et al., 2001; Grefkes and Fink, 2005]. Previous studies have shown that the human PPC/IPS contain polymodal areas [Bremmer et al., 2001], and representations of space that may combine information between different sensory and motor systems. Specifically, in monkey's area VIP, visual and auditory neurons receptive fields substantially overlap, suggesting a role of this region in the supramodal representation of external space [Gross and Graziano, 1995; Schlack et al., 2005; see also Macaluso et al., 2003 for related fMRI results].

Hence, in PPC we found that the BOLD signal was enhanced according to the side of the VE (Fig. 2B), was modulated by AV spatial interactions (Fig. 3B), and was independent of oculomotor behaviour (right panels, Figs. 2–3B). Although partially overlapping in the superior parietal gyrus, the effects of 'side of the VE' and AV spatial interactions appeared to primarily affect different regions of the IPS: more lateral and anterior the former, more ventral and posterior the latter (possibly corresponding to LIP/VIP, see above). The involvement of multiple sub-regions within PPC would be in agreement with the proposal that the parietal cortex computes 'priority maps' (i.e., modality-independent representations of the environment that combine bottom-up and top-down signals) via neurons distributed in adjacent areas of the PPC [Ptak, 2012].

A more parsimonious account of the role of the PPC/IPS in this study would implicate the mechanisms of spatial orienting and spatial attention common between vision and audition [Farah et al., 1989], without implying any integration between the two modalities. As noted above, there was no significant effect of audition on overt orienting behaviour, that is, changes of gaze position could be reliably predicted based on the VE only (Fig. 1D). In light of this, we propose a two-stage mechanism for the processing of the AV events in this study. Accordingly, the VE triggered stimulus-driven and top-down processes that attracted spatial attention towards one hemifield. As a consequence of this, the AE fell (congruent trials) or did not fall (incongruent trials) within the current focus of attention, thus receiving (or not) sufficient processing/attentional resources to modulate activity in the parietal cortex. In this view, the allocation of visual spatial attention would boost the processing of the bottom-up auditory input, as previously suggested in the context of crossmodal spatial effects of audition on visual occipital cortex [McDonald et al., 2003; Romei et al., 2009; Störmer et al., 2009].

Auditory Salience Modulates Activity in the STC

Our imaging results showed that the trial-by-trial variation of auditory salience affected activity in the STC, irrespective of visual condition, sound side and orienting mode. The region modulated by auditory salience was circumscribed within the area activated by the overall effect of sound ('sound vs. no sound' trials, Fig. 2A).

The modulatory influence of auditory salience was located in the STC immediately posterior to the Heschl's gyrus (i.e., in the planum temporale), thus mainly affecting non-primary auditory cortex. Electrophysiological and functional imaging studies converge in showing that a functional dichotomy exists within the auditory cortex, whereby activity in primary areas is more related to stimulus-driven features, whereas activity in non-primary areas is more subjected to task-related attentional modulations [Okamoto et al., 2011; Petkov et al., 2004; Woods et al., 2009]. However, other studies reported that both stimulus-driven and top-down attentional modulations can affect activity in primary [Chait et al., 2010; Sussman et al., 2002] and non-primary auditory cortex [Ahveninen et al., 2011; Grady et al., 1997].

Several previous studies also reported modulation of activity in the auditory cortex when manipulating 'sensory' aspects of sounds, including stimulus presentation rate [Noesselt et al., 2003], automatic detection of acoustic changes [Schönwiesner et al., 2007] and figure-ground segregation [Teki et al., 2011]. These studies considered how changes of one specific stimulus-feature affected brain activity [Leaver and Rauschecker, 2010; Lewis et al., 2012]. Asal maps take into account changes along multiple stimulus dimensions (i.e., intensity, frequency and time). Nonetheless, each feature contributes to the final saliency map and, indeed, a control analysis using sound amplitude rather than saliency also revealed a significant

modulation of activity in the auditory cortex. The relationship between single features and the overall saliency map is complex, and an extensive discussion of this issue would be beyond the scope of this study (for a detailed treatment, see Bordier et al., 2013). However, we should emphasise that sound amplitude should not be regarded as an index of the auditory strength equivalent to auditory salience. Indeed, in this study sound amplitude did not modulate activity in PPC and, previously, we showed that auditory salience can explain activity in the superior temporal gyrus over and above any combination of low-level auditory features [Bordier et al., 2013]. Rather, the saliency model is thought to capture stimulus-driven aspects of auditory attention [Kalinli and Narayanan, 2009; Kayser et al., 2005], linking our findings in the auditory cortex with mechanisms of selection in complex acoustic environments.

CONCLUSIONS

This study investigated spatial orienting in ecologically valid AV settings, examining the BOLD correlates of stimulus-driven signals and their efficacy for overt and covert orienting. The imaging results showed that: (i) stimulus-driven signals modulate activity in visual and auditory sensory cortices; (ii) extrastriate visual cortex (primarily MT+) is modulated by both stimulus-driven signals and the efficacy of VEs in triggering spatial orienting; (iii) auditory salience interacts with AV spatial alignment in the PPC and (iv) these effects are independent of the overt or covert mode of orienting. We conclude that, in naturalistic conditions entailing high levels of sensory competition, activity in sensory areas reflects both stimulus-driven signals and their efficacy for spatial orienting, and that the PPC combines spatial information from the visual and the auditory modality.

ACKNOWLEDGMENTS

The authors thank Sara Pernigotti and Davide Stramaccia for helping them with stimulus recording and editing. The authors declare no conflict of interest, both in general and in relation to this study.

REFERENCES

Ahveninen J, Hämäläinen M, Jääskeläinen IP, Ahlfors SP, Huang S, Lin FH, Raij T, Sams M, Vasios CE, Belliveau JW (2011): Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. Proc Natl Acad Sci USA 108:4182–4187.

Amlôt R, Walker R, Driver J, Spence C (2003): Multimodal visual-somatosensory integration in saccade generation. Neuropsychologia 41:1–15.

Andersen RA, Essick GK, and Siegel RM (1985): Encoding of spatial location by posterior parietal neurons. Science 230:456–458.

Arndt PA, Colonius H (2003): Two stages in crossmodal saccadic integration: Evidence from a visual-auditory focused attention task. Exp Brain Res 150:417–426.

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

Balan PF, Gottlieb J (2006): Integration of exogenous input into a dynamic saliency map revealed by perturbing attention. J Neurosci 26:9239–9249.

Beck DM, Kastner S (2007): Stimulus similarity modulates competitive interactions in human visual cortex. J Vis 7:19.1–12.

Bisley JW, Goldberg ME (2003): Neuronal activity in the lateral intraparietal area and spatial attention. Science 299:81–86.

Bogler C, Bode S, Haynes JD (2011): Decoding successive computational stages of salience processing. Curr Biol 21:1667–1671.

Bordier C, Puja F, Macaluso E (2013): Sensory processing during viewing of cinematographic material: Computational modeling and functional neuroimaging. Neuroimage 67:213–226.

Bremmer F, Ilg UJ, Thiele A, Distler C, Hoffmann KP (1997): Eye position effects in monkey cortex. I. Visual and pursuit-related activity in extrastriate areas MT and MST. J Neurophysiol 77:944–961.

Bremmer F, Schlack A, Shah NJ, Zafiris O, Kubischik M, Hoffmann K, Zilles K, Fink GR (2001): Polymodal motion processing in posterior parietal and premotor cortex: A human fMRI study strongly implies equivalencies between humans and monkeys. Neuron 29:287–296.

Brotchie PR, Andersen RA, Snyder LH, Goodman SJ (1995): Head position signals used by parietal neurons to encode locations of visual stimuli. Nature 375:232–235.

Buschman TJ, Miller EK (2007): Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315:1860–1862.

Carmi R, Itti L (2006): Visual causes versus correlates of attentional selection in dynamic scenes. Vision Res 46:4333–4345.

Chait M, de Cheveigné A, Poeppel D, Simon JZ (2010): Neural dynamics of attending and ignoring in human auditory cortex. Neuropsychologia 48:3262–3271.

Colby CL, Goldberg ME (1999): Space and attention in parietal cortex. Annu Rev Neurosci 22:319–349.

Constantinidis C (2006): Posterior parietal mechanisms of visual attention. Rev Neurosci 17:415–427.

Constantinidis C, Steinmetz MA. (2005): Posterior parietal cortex automatically encodes the location of salient stimuli. J Neurosci 25:233–238.

Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1990): Attentional modulation of neural processing of shape, color, and velocity in humans. Science 248:1556–1559.

Corbetta M, Akbudak E, Conturo TE, Snyder AZ, Ollinger JM, Drury HA, Linenweber MR, Petersen SE, Raichle ME, Van Essen DC, Shulman GL (1998): A common network of functional areas for attention and eye movements. Neuron 21:761–773.

Corneil BD, Munoz DP (1996): The influence of auditory and visual distractors on human orienting gaze shifts. J Neurosci 16:8193–8207.

Corneil BD, Van Wanrooij M, Munoz DP, Van Opstal AJ (2002): Auditory-visual interactions subserving goal-directed saccades in a complex scene. J Neurophysiol 88:438–454.

Desimone R, Duncan J (1995): Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222.

Duhamel JR, Bremmer F, Ben Hamed S, Graf W (1997): Spatial invariance of visual receptive fields in parietal cortex neurons. Nature 389:845–848.

Duvernoy HM. 1991. The Human Brain: Surface, Three-Dimensional Sectional Anatomy and MRI. Wien: Springer.

Einhäuser W, Rutishauser U, Koch C (2008): Task-demands can immediately reverse the effects of sensory-driven salience in complex visual stimuli. J Vis 8:2.1–19.

Elazary L, Itti L (2008): Interesting objects are visually salient. J. Vis 3:1–15.

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

Farah MJ, Wong AB, Monheit MA, Morrow LA (1989): Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal? Neuropsychologia 27:461–470.

Frens MA, Van Opstal AJ, Van der Willigen RF (1995): Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. Percept Psychophys 57:802–816.

Friston KJ, Glaser DE, Henson RN, Kiebel S, Phillips C, Ashburner J (2002): Classical and Bayesian inference in neuroimaging: Applications. Neuroimage 16:484–512.

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

Grady CL, Van Meter JW, Maisog JM, Pietrini P, Krasuski J, Rauschecker JP (1997): Attention-related modulation of activity in primary and secondary auditory cortex. Neuroreport 8:2511–2516.

Grefkes C, Fink GR (2005): The functional organization of the intraparietal sulcus in humans and monkeys. J Anat 207:3–17.

Gross CG, Graziano MS (1995): Multiple representations of space in the brain. Neuroscientist 1:43–50.

Grunewald A, Linden JF, Andersen RA (1999): Responses to auditory stimuli in macaque lateral intraparietal area. I. Effects of training. J Neurophysiol 82:330–342.

Helbig HB, Ernst MO, Ricciardi E, Pietrini P, Thielscher A, Mayer KM, Schultz J, Noppeney U (2012): The neural mechanisms of reliability weighted integration of shape information from vision and touch. Neuroimage 60:1063–1072.

Itti L, Koch C (2001): Computational modelling of visual attention. Nat Rev Neurosci 2:194–203.

Itti L, Koch C, Niebur E (1998): A model of salience-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal Mach Intell 20:1254–1259.

Jack AI, Patel GH, Astafiev SV, Snyder AZ, Akbudak E, Shulman GL, Corbetta M (2007): Changing human visual field organization from early visual to extra-occipital cortex. PLoS One 2:e452.

Kalinli O, Narayanan S (2009): Prominence detection using auditory attention cues and task-dependent high level information. IEEE Trans Audio Speech Lang Process 17:1009–1024.

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999): Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22:751–761.

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

Koelewijn T, Bronkhorst A, Theeuwes J (2010): Attention and the multiple stages of multisensory integration: A review of audiovisual studies. Acta Psychol 134:372–384.

Koyama M, Hasegawa I, Osada T, Adachi Y, Nakahara K, Miyashita Y (2004): Functional magnetic resonance imaging of macaque monkeys performing visually guided saccade tasks: Comparison of cortical eye fields with humans. Neuron 41:795–807.

Leaver AM, Rauschecker JP (2010): Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. J Neurosci 30:7604–7612.

Leopold DA, Logothetis NK (1998): Microsaccades differentially modulate neural activity in the striate and extrastriate visual cortex. Exp Brain Res 123:341–345.

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

Lewis JW, Talkington WJ, Tallaksen KC, Frum CA (2012): Auditory object salience: Human cortical processing of non-biological action sounds and their acoustic signal attributes. Front Syst Neurosci 6:27.

Linden JF, Grunewald A, Andersen RA (1999): Responses to auditory stimuli in macaque lateral intraparietal area. II. Behavioral modulation. J Neurophysiol 82:343–358.

Macaluso E (2010): Orienting of spatial attention and the interplay between the senses. Cortex 46:282–297.

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

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.

Mazzoni P, Bracewell RM, Barash S, Andersen RA (1996): Spatially tuned auditory responses in area LIP of macaques performing delayed memory saccades to acoustic targets. J Neurophysiol 75:1233–1241.

McDonald JJ, Ward LM (2000): Involuntary listening aids seeing: Evidence from human electrophysiology. Psychol Sci 11:167–171.

McDonald JJ, Teder-Sälejärvi WA, Hillyard SA (2000): Involuntary orienting to sound improves visual perception. Nature 407:906–908.

McDonald JJ, Teder-Sälejärvi WA, Di Russo F, Hillyard SA (2003): Neural substrates of perceptual enhancement by cross-modal spatial attention. J Cogn Neurosci 15:10–19.

McGurk H, MacDonald J (1976): Hearing lips and seeing voices. Nature 265:746–748.

McMains SA, Kastner S (2010): Defining the units of competition: Influences of perceptual organization on competitive interactions in human visual cortex. J Cogn Neurosci 22:2417–2426.

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

Nardo D, Santangelo V, Macaluso E (2011): Stimulus-driven orienting of visuo-spatial attention in complex dynamic environments. Neuron 69:1015–1028.

Nobre AC, Gitelman DR, Dias EC, Mesulam MM (2000): Covert visual spatial orienting and saccades: Overlapping neural systems. Neuroimage 11:210–216.

Noesselt T, Shah NJ, Jäncke L (2003): Top-down and bottom-up modulation of language related areas—An fMRI study. Biomed Chromatogr Neurosci 4:13.

Nuthmann A, Henderson JM (2010): Object-based attentional selection in scene viewing. J Vis 10:20.

Okamoto H, Stracke H, Bermudez P, Pantev C (2011): Sound processing hierarchy within human auditory cortex. J Cogn Neurosci 23:1855–1863.

Onat S, Libertus K, König P (2007): Integrating audiovisual information for the control of overt attention. J Vis 7:11.1–11.16.

Penny W, Holmes A (2004): Random effects analysis. In: Frackowiak RS, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, Ashburner J, Penny W, editors. Human Brain Function II, 2nd ed. San Diego, CA: Elsevier. pp843–851.

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

Ptak R (2012): The frontoparietal attention network of the human brain: Action, salience, and a priority map of the environment. Neuroscientist 18:502–515.

Recanzone GH, Wurtz RH (2000): Effects of attention on MT and MST neuronal activity during pursuit initiation. J Neurophysiol 83:777–790.

Reddy L, Kanwisher NG, VanRullen R (2009): Attention and biased competition in multi-voxel object representations. Proc Natl Acad Sci USA 106:21447–21452.

Rizzolatti G, Riggio L, Dascola I, Umiltá C (1987): Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. Neuropsychologia 25:31–40.

Romei V, Murray MM, Cappe C, Thut G (2009): Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. Curr Biol 19:1799–1805.

Santangelo V, Olivetti Belardinelli M, Spence C, Macaluso E (2009): Interactions between voluntary and stimulus-driven spatial attention mechanisms across sensory modalities. J Cogn Neurosci 21:2384–2397.

Schiller PH (1993): The effects of V4 and middle temporal (MT) area lesions on visual performance in the rhesus monkey. Vis Neurosci 10:717–746.

Schlack A, Sterbing-D'Angelo SJ, Hartung K, Hoffmann KP, Bremmer F (2005): Multisensory space representations in the macaque ventral intraparietal area. J Neurosci 25:4616–4625.

Schönwiesner M, Novitski N, Pakarinen S, Carlson S, Tervaniemi M, Näätänen R (2007): Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. J Neurophysiol 97:2075–2082.

Snyder LH, Grieve KL, Brotchie P, Andersen RA (1998): Separate body- and world-referenced representations of visual space in parietal cortex. Nature 394:887–891.

Spence C, Driver J (1997): Audiovisual links in exogenous covert spatial orienting. Percept Psychophys 59:1–22.

Spence C, Nicholls ME, Gillespie N, Driver J (1998): Cross-modal links in exogenous covert spatial orienting between touch, audition and vision. Percept Psychophys 60:544–557.

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

Stein BE, Meredith MA, Huneycutt WS, McDade L (1989): Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. J Cogn Neurosci 1:12–24.

Störmer VS, McDonald JJ, Hillyard SA (2009): Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. Proc Natl Acad Sci USA 106:22456–22461.

Sussman E, Winkler I, Huotilainen M, Ritter W, Näätänen R (2002): Top-down effects can modify the initially stimulus-driven auditory organization. Brain Res Cogn Brain Res 13:393–405.

Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010): The multifaceted interplay between attention and multisensory integration. Trends Cogn Sci 14:400–410.

Teki S, Chait M, Kumar S, von Kriegstein K, Griffiths TD (2011): Brain bases for auditory stimulus-driven figure-ground segregation. J Neurosci 31:164–171.

Tseng PH, Carmi R, Cameron IG, Munoz DP, Itti L (2009): Quantifying center bias of observers in free viewing of dynamic natural scenes. J Vis 9:4–16.

Woods DL, Stecker GC, Rinne T, Herron TJ, Cate AD, Yund EW, Liao I, Kang X (2009): Functional maps of human auditory cortex: Effects of acoustic features and attention. PLoS One 4:e5183.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996): A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp 4:58–73.