Europe PMC Funders Group
Author Manuscript
Neuroimage. Author manuscript; available in PMC 2013 November 25.

Published in final edited form as:
Neuroimage. 2013 February 15; 67: . doi:10.1016/j.neuroimage.2012.11.031.

Europe PMC Funders Author Manuscripts

Sensory processing during viewing of cinematographic material: Computational modeling and functional neuroimaging

Cecile Bordier[*], **Francesco Puja**, and **Emiliano Macaluso**
Neuroimaging Laboratory, IRCCS, Santa Lucia Foundation, Via Ardeatina 306, Rome 00179, Italy

Abstract

The investigation of brain activity using naturalistic, ecologically-valid stimuli is becoming an important challenge for neuroscience research. Several approaches have been proposed, primarily relying on data-driven methods (e.g. independent component analysis, ICA). However, data-driven methods often require some post-hoc interpretation of the imaging results to draw inferences about the underlying sensory, motor or cognitive functions. Here, we propose using a biologically-plausible computational model to extract (multi-)sensory stimulus statistics that can be used for standard hypothesis-driven analyses (general linear model, GLM). We ran two separate fMRI experiments, which both involved subjects watching an episode of a TV-series. In Exp 1, we manipulated the presentation by switching on-and-off color, motion and/or sound at variable intervals, whereas in Exp 2, the video was played in the original version, with all the consequent continuous changes of the different sensory features intact. Both for vision and audition, we extracted stimulus statistics corresponding to spatial and temporal discontinuities of low-level features, as well as a combined measure related to the overall stimulus saliency. Results showed that activity in occipital visual cortex and the superior temporal auditory cortex co-varied with changes of low-level features. Visual saliency was found to further boost activity in extra-striate visual cortex plus posterior parietal cortex, while auditory saliency was found to enhance activity in the superior temporal cortex. Data-driven ICA analyses of the same datasets also identified “sensory” networks comprising visual and auditory areas, but without providing specific information about the possible underlying processes, e.g., these processes could relate to modality, stimulus features and/or saliency. We conclude that the combination of computational modeling and GLM enables the tracking of the impact of bottom–up signals on brain activity during viewing of complex and dynamic multisensory stimuli, beyond the capability of purely data-driven approaches.

Keywords

Data-driven; Saliency; Multi-sensory; Cinematographic material; Biologically-inspired vision and audition

Introduction

Functional imaging has been used extensively to non-invasively map sensory, motor and cognitive functions in humans. Nonetheless, so far the vast majority of studies have

© 2012 Elsevier Inc. All rights reserved.

[*]Corresponding author. Fax: +39 06 5150 1213. c.bordier@hsantalucia.it (C. Bordier). .

Appendix A. Supplementary data
Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2012.11.031.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

employed simple and repeated stimuli that are in striking contrast with the unrepeated, complex and dynamic signals that the brain has to process in everyday life. Moreover, the neuronal responses in conventional laboratory conditions, i.e. using artificial stimuli, are weaker than those associated with naturalistic stimuli (Mechler et al., 1998; Yao et al., 2007). Thus, recently, a growing interest has risen around the use of more ecologically-valid stimuli during fMRI (e.g. Hasson et al., 2010).

A central issue with naturalistic approaches is that, unlike standard paradigms, there is no straightforward correspondence between the stimuli presented to the subject and any specific sensory, motor or cognitive function. This makes it difficult to use hypothesis-based analysis methods that involve fitting BOLD data with predictors representing specific experimental conditions or processes (general linear model; see Friston et al., 2003). Indeed, many previous studies using complex and dynamic stimuli (e.g. cinematographic material) have resorted to data-driven approaches that do not require any such “a priori” coding.

One of these approaches relies on multivariate analysis based on independent component analysis (ICA) (Bartels and Zeki, 2005; Calhoun et al., 2001a, 2001b; McKeown et al., 1998). ICA performs a blind separation of independent sources from the complex mixture of signal and noise resulting from many different sources. This method does not require any “a priori” specification of the possible causes of the responses (i.e. predictors in a standard GLM analysis) and even no specification of the shape of the hemodynamic response function (HRF). Instead, the method is based on the intrinsic structure of the data. ICA aims to extract a number of unknown sources of signal that are mutually and statistically independent in space or time. Friston (1998) showed the relevance of this idea to biological time-series.

Inter-subject correlation (ISC) analyses is another data-driven approach that has been recently introduced to investigate brain activity associated with the processing of complex stimuli (Hasson et al., 2004; Sui et al., 2012). ISC analyses are based on the idea that presenting the same complex and dynamic sensory input to different subjects will generate the same pattern of BOLD activity in the brain areas processing the stimuli. Therefore, these areas can be identified by testing for correlated BOLD time-courses between subjects (“synchronization”). This approach has now been employed with a variety of stimuli and tasks (e.g., palm trees task, Seghier and Price, 2009; face processing, Lessa et al., 2011; story comprehension, Lerner et al., 2011; movie watching, Kauppi et al., 2010; representation of action-schemas, Hanson et al., 2009).

However, these data-driven methods have the intrinsic limitation that they require some post-hoc interpretation regarding the processes that generate ICA components (e.g. motion in MT, Bartels and Zeki, 2005) or patterns of inter-subject synchronization (e.g. faces in FFA, Hasson et al., 2004). Here, we propose using an alternative approach where (multi-) sensory stimulus statistics are first extracted via a biologically-plausible computational model and are then used for hypothesis-driven analyses (see also Nardo et al., 2011; Bartels et al., 2008).

One of the most successful biologically-plausible computational model of sensory bottom–up processing consists in the computation of “saliency maps” from complex, naturalistic images (Itti and Koch, 2001; Koch and Ullman, 1985). Inspired from the organization of the visual system, saliency maps are based on the extraction of local discontinuities in intensity, color, orientation, motion and flicker (i.e. feature-specific maps). Saliency maps are then computed as a combination of these feature maps. Saliency maps are thought to well-characterize the spatial distribution of bottom–up signals and have been found to predict

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

sequences of fixations during viewing of naturalistic pictures (Parkhurst et al., 2002) and video-clips (Itti, 2005).

While used primarily in eye-movement studies, saliency has been also considered in electrophysiological studies in monkeys (Gottlieb, 2007; Gottlieb et al., 1998; Thompson et al., 2005) and fMRI in humans (Bogler et al., 2011; Nardo et al., 2011). These studies have indicated that visual saliency is represented in visual areas (Li, 2002; VanRullen, 2003), and can influence activity in higher-order parietal/frontal areas as a function of specific attentional operations (i.e. efficacy of the bottom–up signals for spatial orienting, Nardo et al., 2011; winner-take-all mechanism of attentional selection, Bogler et al., 2011). However, studies employing saliency models took into account only the final “saliency map”, with little consideration of the possible contribution of the feature-specific maps (Parkhurst et al., 2002; Itti, 2005; Bartels et al., 2008; but see Itti et al., 1998; Liu et al., 2011). Here we used concurrently *both* saliency and feature maps to investigate the impact of bottom–up sensory signals on brain activity during viewing of complex stimuli.

More recently, the use of computational models of saliency has been extended to the analysis of complex signals in modalities other than vision. Using an approach analogous to the original model for vision (Itti et al., 1998; Koch and Ullman, 1985), Kayser and colleagues proposed a method to compute the saliency of complex, naturalistic auditory stimuli (Kayser et al., 2005; see also Altmann et al., 2008; Kalinli and Narayanan, 2007). Again, saliency maps are constructed combining feature maps that, for audition, extract discontinuities in intensity, frequency and time. Here, for the first time, we propose using auditory saliency – as well as auditory features – to investigate brain activity recorded while volunteers were presented with complex auditory stimuli.

Accordingly, the goal of this study was to use computationally-derived indexes of visual and auditory bottom–up signals (i.e. saliency and features) to assess brain activity during the presentation of complex audio-visual stimuli (cinematographic material); and to evaluate this with respect to a conventional block/event condition-based analysis (cf. Exp 1) and a fully data-driven approach (ICA).

Methods

Participants

This study included two experiments. Eight healthy volunteers took part in the first experiment (Exp1; 2 males, range: 21–24, mean age: 23) and seven different volunteers in the second experiment (Exp2a and Exp2b; 1 male, range: 21–23, mean age: 22). All volunteers were Italian speaking, had normal or corrected-to-normal vision, and did not report any neurological impairment. After having received instructions, all participants gave their written consent. The study was approved by the independent Ethics Committee of the Santa Lucia Foundation (Scientific Institute for Research Hospitalization and Health Care).

Stimuli

In Experiment 1, subjects were asked to view half of an episode of the TV-series “24” (21 min 40 s). For this experiment, we manipulated the original video by switching on-and-off color, motion and/or sound at variable intervals. This provided us with a dataset containing a known correspondence between the stimuli and sound/feature-related sensory processes, which enabled us to perform conventional “condition-based” analyses (see below). The three sensory streams (sound: on/off; motion: motion/static; color: color/black-white) were switched on-and-off in blocks with durations ranging between 8.3 and 43.7 s. The time course of the on-off sequences was independent in the three streams, thus including all

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

possible combinations of the three sound/features input. All subjects were presented with the same version of the modified video (i.e. same sequences of on/off-sets).

In Experiment 2, subjects were presented with the original version of the video, without any manipulation of sound, motion or color. Thus, unlike Exp1, now there was a continuous and unknown change of the different sensory input over time. The experiment was divided into two scanning sessions that consisted of the presentation of the first half (Exp2a: 21 min 40 s) and the second half (Exp2b: 20 min 07 s) of the same “24” episode used in Exp1. None of the volunteers of Exp2 had taken part in Exp1, so they watched the TV episode for the first time. The fMRI data analyses were carried out separately for Exp2a and Exp2b, with the aim of assessing the reproducibility of the results.

For both experiments, visual stimuli were back-projected on a semi-opaque screen at the back of the magnet. Participants viewed the screen through a mirror located above their eyes. The size of the visual display was approximately (24°×16° visual angle); the sound was presented via MRI-compatible headphones.

Parameterization of visual/auditory features and saliency

Visual saliency model—The visual saliency model allows identification of salient locations within a given image (Itti et al., 1998; Koch and Ullman, 1985). Our current implementation was modified from the software available at: http://www.saliencytoolbox.net. This considers static visual features only (intensity, color and orientation), while here we added motion and flicker contrasts for the analysis of our dynamic visual stimuli (see Itti and Pighin, 2003). The step-by-step detailed description of the procedures and the specific parameters used to compute the visual saliency maps are reported in the Supplementary materials (Table S1).

Modeling consists of decomposing the input image/s into a set of distinct “channels” using linear filters tuned to specific stimulus dimensions. This decomposition is performed at several levels, extracting contrasts at different spatial scales (Gaussian pyramids; Greenspan et al., 1994). The scales were created using pyramids with 9 levels. Together with intensity, color and orientation, here flicker was computed from the absolute difference between the intensity of the current frame and that of the previous frame. Motion was computed from spatially-shifted differences between the intensity pyramids from the current and the previous frame (Itti and Pighin, 2003; Reichardt, 1987). Center-surround interactions are then implemented as differences between fine and coarse scales of the pyramids, here using a set of 6 cross-scale subtractions for each channel of each feature (Itti and Pighin, 2003; Walther and Koch, 2006; see Table S1, for details). This produced a total of 72 maps, with 6 maps for the intensity feature (1 channel: on/off contrast); 12 maps for the color feature (2 channels: red/green and blue/yellow contrasts); 24 maps for the orientation feature (4 channels: contrasts at 0°, 45°, 90° and 135°); 24 maps for the motion feature (4 channels: contrasts at 0°, 45°, 90° and 135°); and 6 maps for the flicker feature (1 channel: on/off contrast).

Iterative non-linear filtering was used to simulate the competition between salient locations (4 iterations). Each iteration consists of convolving the maps with a 2D Difference of Gaussian (DoG; Itti and Koch, 2000, see also Table S1). This imitates self-excitation and inhibition induced by neighboring peaks, thus implementing competitive interactions between different spatial locations in the map. After this, for each of the 5 features, the maps were summed across scales (at equivalent pixel locations) to generate the “conspicuity maps”. Each feature was given equal weight (=1), taking into account the number of maps available for that feature (cf. above and Table S1). These un-normalized conspicuity maps were used to compute feature-specific regressors for the fMRI analyses (see section below).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

To generate the final saliency map, each conspicuity map was normalized again with the DoG filter, the five normalized conspicuity maps were summed across equivalent pixel locations, and normalized once more with the DoG filter (Itti and Koch, 2000; Itti et al., 1998; see Table S1). The final spatial resolution of the maps was 33 by 45 pixels.

In sum, for each frame of the movie we generated one saliency map and five feature-specific maps (intensity, color, orientation, motion and flicker) that were then used to construct saliency and feature-specific regressors for the fMRI analyses, as we detail in the section “Visual and auditory regressors” below.

Auditory saliency model—The computation of auditory saliency was conceptually analogous to the model used for visual saliency (see Kayser et al., 2005; and Table S2, for details). Sound segments from the movie soundtrack (each with 2.08 s duration, see Fig. 1) were preprocessed using a Fast Fourier analysis (37 ms windows, 95% overlap, frequency band 100 Hz to 10 kHz and decision regions of 20 Hz, Shamma, 2001). This resulted in a time-by-frequency image of 136-by-62 pixels, for each 2.08 s auditory segment. These images were then analyzed using feature detectors on different scales (i.e. Gaussian pyramids with 8 levels; see Table S2). The features extracted were intensity, frequency contrast, temporal contrast (cf. Kayser et al., 2005), plus orientation. The orientation contrast mimics the dynamics of auditory neuronal response to moving ripples in the primary auditory cortex (Kalinli and Narayanan, 2007). Center-surround differences were then computed with a set of 4 cross-scale subtractions (see Table S2). The cross-scale subtraction was performed for each channel of each auditory feature (i.e. 1 channel for intensity, frequency, and temporal, and 2 channels for orientation) producing a total of 20 auditory maps.

All 20 maps were normalized using DoG filters (Itti and Koch, 2001; Kalinli and Narayanan, 2007; see Table S2, for the parameters details) and combined into conspicuity maps using across-scale additions at equivalent pixel locations (weight=1, for each of the four features). These maps were used to compute the feature-specific auditory regressors for the fMRI analyses (see section below). Finally, the DoG normalization was performed again on the conspicuity maps and the results were summed to obtain the final auditory saliency map (see Table S2). The maxima of the saliency map define salient points in the 2-D time-frequency auditory spectrum.

The extraction of the auditory conspicuity/saliency maps was performed considering 2.08 s segments (corresponding to the repetition time of the fMRI), because center-surround and normalization steps would be inappropriate at temporal scales of several tens of minutes (i.e. total soundtrack duration). Nonetheless, we enabled interactions between sounds in different segments by computing conspicuity/saliency maps in segments with 50% overlaps and then averaging values over the overlapping time-frames (see Fig. 1). This procedure generated one saliency map and four feature-specific maps (intensity, temporal, frequency and orientation contrast) for each segment (duration=1 TR) and each channels (right and left). Left and right channels were then averaged before constructing the regressors for the fMRI analyses, see section below.

Visual and auditory regressors—The methods described above provided us with saliency maps and feature-specific maps (i.e. the un-normalized conspicuity maps) for vision and audition. The next step was to convert these high-dimensional matrices (vision: vertical position×horizontal position×time; audition: frequency×time) into regressors for the SPM design matrix, with a single value for each fMRI volume (see Fig. 1, for a schematic representation of this procedure).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

For the visual features, we estimated the mean value of each un-normalized conspicuity map (i.e. over the vertical and horizontal spatial dimensions) and averaged this to the fMRI repetition time (i.e. from 25 points per second, to 1 point per TR=2.08 s). For the visual saliency maps, which contain discrete clusters (e.g. see Fig. 1), we computed the mean of each cluster, and then averaged these values across all clusters and re-sampled to the fMRI repetition time.

For the auditory features, we averaged the values of the unnormalized conspicuity maps over frequency and time within each segment. Since segments were chosen to have the same duration as the fMRI repetition time, no further re-sampling over time was required. For the auditory saliency, we extracted the maximum value over frequency for each time point of the map and then averaged over time. Note that we used different approaches for visual saliency (mean over clusters) and auditory saliency (maximum over frequencies), because the two corresponding saliency maps are markedly different (cf. Fig. 1, vision on the left and audition on the right; with most pixels equal to zero in the auditory saliency map).

Finally, all vectors (1 saliency and 5 features, for vision; 1 saliency and 4 features, for audition) were convolved with the SPM8 hemodynamic response function (HRF) in order to generate the final 11 regressors used as predictors in the SPM design matrix.

FMRI acquisition and pre-processing

Images were acquired with a Siemens Allegra (Siemens Medical System Erlangen, Germany) 3 T scanner equipped for echo-planar imaging (EPI). A transmit-receive quadrature birdcage head coil was used. Functional imaging data were acquired using gradient-echo echo-planar imaging (TR/TE=2.08 s/30 ms, flip angle=70 °, matrix 64×64, voxel size=3×3 mm in-plane, slice thickness=2.5 mm; 50% distance factor), with 32 contiguous transverse slices covering the entirety of the cerebral cortex.

The data were pre-processed using SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK). After discarding the initial volumes (4 for Exp 2a/b; and 24 for Exp1, because of hardware stabilization problems), the remaining volumes (Exp1: 611; Exp2a: 627; Exp2b: 583) were slice-timed, head-motion realigned and normalized to the standard MNI EPI template space (voxel-size re-sampled to 3×3×3 mm^3). Finally, the data were spatially smoothed with a 8×8×8 mm^3 full-width at half-maximum Gaussian kernel.

fMRI analyses

fMRI data assessment was carried out using hypothesis-based analyses (general linear models, GLMs) and a data-driven approach (independent component analysis, ICA). Both GLMs and ICA analyses used a fixed-effects approach, because here we did not seek to generalize inference to the population. Rather, our aim was to evaluate and compare different methodological approaches for the investigation of brain activity during stimulation with complex and dynamic stimuli. Thus, here statistical inference concerns only the groups of subjects who took part in the current experiments. However, note that we performed analogous analyses with three independent datasets (Exp1, 2a and 2b; including two different groups of subjects) seeking to confirm the reproducibility of our results.

Condition-based GLM analysis (Exp1 only)—Exp1 entailed the experimental manipulation of sound, motion and color on/off-sets. This enabled us to perform a standard “condition-based” GLM analysis (SPM8). The statistical model included 6 predictors of interest; plus subject-specific realignment-parameters and session/subject constants, as effects of no interest. For each condition (sound, motion, color), the predictors of interest included one regressor for the sustained block-effect (variable duration=8.3–43.7 s) and one

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

for the transient block-onset (duration=0), convolved with the canonical hemodynamic response function (HRF). The time-series were high-pass filtered at 128 s and pre-whitened by means of autoregressive model AR(1). Statistical significance was assessed using three F-contrasts testing for the combined effect of block- and event-predictors, separately for sound, motion and color. The threshold was set to voxel-level p-FWE=0.05, corrected for multiple comparisons considering the whole brain as the volume of interest (see also legend of Fig. 2).

Features and saliency-based GLM analyses—The main aim of the current study was to assess whether parameters derived from computational analyses of complex and dynamic audio-visual stimuli predict brain activity measured during viewing of these complex stimuli.

For all three datasets (Exp1, 2a and 2b), we constructed GLMs including regressors derived from visual and auditory conspicuity maps (feature-predictors); plus regressors derived from visual and auditory saliency maps (saliency-predictors). Visual features included 5 regressors related to color, intensity, orientation, motion and flicker contrasts; auditory features included 4 regressors related to intensity, frequency, temporal and orientation contrasts (see section above for details). Accordingly, each GLM model comprised 11 regressors of interest, plus subject-specific realignment-parameters and the session constant as effects of no interest. The data were high-pass filtered at 128 s and pre-whitened by means of autoregressive model AR(1).

Because of the high correlation between the GLM regressors (cf. also Bartels et al., 2008), the significance of the feature-predictors was assessed using F-contrasts testing for the combined effect of the 5 (visual) or the 4 (auditory) features. The highest correlations concerned the within-modality feature-regressors, in some cases with r-values>0.9. The correlation between feature- and saliency-regressors had a lower range (−0.19 to 0.64). The between-modalities correlations were relatively low compared with the within-modality correlations (the highest r-value was =0.46; between visual motion and auditory frequency contrast, in Exp2b). It should be noted that the high correlations between regressors will affect the specific values of the GLM parameter estimates, but not the significance of the fit of the model (Andrade et al., 1999). This is the reason why we used F-tests jointly assessing the significance of all feature-regressors within each modality, and do not report any statistics about single features. The statistical threshold for the F-tests was set to voxel-level p-FWE=0.05, corrected for multiple comparisons considering the whole brain as the volume of interest. For each dataset, the corresponding minimum voxel-level F-value is reported in the legend of Fig. 3.

For the saliency predictors, which had lower correlations with the other regressors in the model, we used t-tests looking for areas where activity increased with increasing visual and/or auditory saliency. It should be noted that because the GLMs included both features and saliency regressors in the same model, these t-tests will highlight brain activation associated with saliency over and above any feature-related effect: i.e. the saliency-regressors will fit variance that cannot be accounted for by any combination of the feature-regressors. For these more subtle tests, the threshold was set at p-FWE=0.05 corrected for multiple comparisons, but now at the cluster-level. The cluster-level statistics combines peak-high and cluster-size and requires an additional voxel-level threshold to define the cluster-size. This additional voxel-level threshold was set at p-unc.=0.001, corresponding to a minimum voxel-level T-value=3.09, and the minimum cluster-size was 80 voxels (both for visual and auditory saliency, in all three datasets). Again, the thresholding procedure ensured correction for multiple comparisons at the whole brain level.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Independent component analysis (ICA)—The three pre-processed datasets (Exp1, 2a and 2b) were submitted to the “spatial ICA” Toolbox GIFT (http://icatb.sourceforge.net; Calhoun et al., 2001a). Using the InfoMax ICA algorithm (Bell and Sejnowski, 1995) and the “minimum description length criterion” (Li et al., 2006) we extracted 25 components for Exp1, and 28 components both for Exp2a and for Exp2b.

We seek to assess the consistency of the spatial patterns of the components between datasets by computing a “% overlap” index. For this, we considered all possible combinations (triplets) including one component of each dataset. For each triplet, the “% overlap” index was calculated as the sum of the number of voxels common to the three datasets divided by the number of voxels in the component with fewer voxels. The index is equal to 0 if there is no voxel common to the three datasets; and is equal to 1 (i.e. 100% overlap) if all voxels of the dataset with the “smaller component” (i.e. fewer voxels) are also found in the other two datasets. Following the computation of this index for all possible triplets (n=19600), we considered further only components that belonged to triplets with an index larger than 50%. This allowed us to identify 15 consistent components. Out of these, by visual inspection, 5 were categorized as artifacts (i.e. ridges at the border of the brain, eye-balls, or cerebrospinal fluid). Accordingly, we retained 10 consistent components that we could identify in each of the three datasets (see Fig. 5 and Fig. S1, in the Supplemental material). To facilitate comparisons between the three datasets, consistent components were tentatively labeled as “visual” (2 components), “auditory” (1 component) or “other” (7 components); see Results section for details about these components.

The GIFT Toolbox offers the possibility to realize multi-regressions between the temporal profile of each ICA component and predictors from SPM models. Here, we used the same SPM models as in our GLM analyses (see above). Separately for the three datasets, we considered features and saliency predictors and computed corresponding r-square values for each of the 10 consistent components (see Fig. 6). The r-square values capture the relationship between the temporal profile of each ICA component and changes of the features or saliency over time. Corresponding F-values were computed as follows:

$$F=\frac{\left(R^2/k\right)}{\left(1-R^2\right)/\left(N-k-1\right)}.$$

N number of points/scans in the component

k number of tested regressors.

We report as statistically significant r-square coefficients with associated p-values<0.001 (see Fig. 6).

Results

Condition-based GLM analysis (Exp 1 only)

The results of the GLM analysis based on standard condition-specific regressors (on/off of sound, color and motion) are presented in Fig. 2 and Table 1. Activation maps report the F-statistic jointly testing for sustained (block) and transient (event) activation associated with the three types of input that were manipulated experimentally in Exp1.

The color-related F-contrast revealed activation in medial occipital regions, right superior occipital gyrus and in the right inferior occipital gyrus (Fig. 2A). The latter may correspond

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

to the color sensitive visual area V4, with 77.8% of the small activation cluster (8 voxels) localized in V4 according to the SPM Anatomy Toolbox.

The motion-related contrast highlighted bilateral activation of the occipital cortex and the occipito-temporal junction. This included the entirety of motion sensitive visual area V5/MT (100% of V5/MT in both hemispheres, according to the SPM Anatomy Toolbox). However activation extended well beyond visual occipital cortex, with clusters located in the posterior parietal cortex, temporal and frontal regions plus the insulae (cf. Fig. 2B and Table 1). This extensive pattern of activation can be explained considering that the "motion vs. no-motion" contrast effectively compared watching dynamic visual stimuli – entailing the presentation of many different objects, scenes, events, etc. – versus just looking at a single static picture. Somewhat surprisingly, this contrast also revealed activation of the posterior part of the superior temporal gyrus: an area that responded to auditory stimulation (see section below) and that included parts of the primary auditory cortex in Heschel's gyrus (see Table 4). Since the F-test is non-directional and can identify both positive and negative BOLD responses, we inspected the parameter estimates associated with block and event motion-regressors in the superior temporal gyrus. This revealed a positive effect for the sustained/ block regressor (with the transient-response not different from zero), suggesting crossmodal activation of auditory cortex by visual signals.

The sound-related contrast showed bilateral activation of the superior temporal gyrus, including nearly the whole of the primary auditory cortex (see Fig. 2C and Table 4). Additional clusters of activation comprised the middle and the inferior frontal gyrus, as well as a cluster in medial frontal cortex (see Table 1). Sound-related effects were found also in the occipital cortex (see Fig. 2C, left panel), even though inspection of the parameter estimates revealed that this was driven primarily by negative sustained/block parameters associated with the sound-regressor. This would be consistent with sound-induced deactivation of visual cortex (Laurienti et al., 2002).

Feature-based GLM analysis

Next we turned to the GLM analyses that used predictors based on features and saliency indexes derived from computational analyses of the audio-visual stimuli.

In Exp1, the F-contrast assessing the 5 regressors based on visual features highlighted significant effects within occipital cortex, with clusters comprising the medial surface (i.e. primary visual cortex), superior and inferior occipital gyri and the occipito-temporal junction. The latter overlapped with the motion-related effect found with the condition-based analysis (cf. Figs. 2B and 3A, two leftmost columns). This contrast also reveled effects of visual features in the posterior parietal cortex and, consistent with the condition-based GLM, cross-modally in the superior temporal gyrus (see also Table 4). Compared with condition-based analyses, the visual-features contrast now showed less extensive activation in the frontal and pre-frontal cortex, suggesting greater selectivity of the feature-based GLM that did not require comparing dynamic vs. static visual input (cf. motion contrast in Fig. 2B).

The F-contrast assessing the 4 auditory-features revealed a pattern of activation consistent with the condition-based analysis (cf. Figs. 2C and 3A rightmost panel). Auditory features were found to affect activity in the superior temporal gyrus bilaterally, including the primary auditory cortex (see Table 4). Also the feature-based GLM confirmed that auditory features affected activity in occipital cortex, with significant clusters found in medial and lateral and inferior occipital cortex (see Table 2, Exp1). However, because of the high correlations between features-regressors the interpretation of the corresponding parameter estimates is ambiguous (see Methods section). Hence, unlike for the standard block/event GLM analysis,

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

we cannot conclude whether auditory features increased or decreased activity in visual cortex.

Overall, the computational feature-based approach was able to replicate the results of the more traditional condition-based analyses. However, it should be noted that in Exp1 feature-based regressors entailed on-off transitions similar to the condition-based regressors, because of the experimental manipulation of the videos. Accordingly, we seek to confirm the results of the feature-based analyses in Exp2a and Exp2b, neither of which included any such artificial manipulation.

Concerning the visual features, F-contrasts in Exp2a and Exp2b revealed patterns of activation similar to Exp1 (see Figs. 3B and C, two leftmost columns; and Table 2). These included the expected effects in occipital cortex (striate and extra-striate), plus several clusters in parietal and temporal cortex. Nonetheless, the “cross-talk” between vision and audition (i.e. changes of visual features showing co-variation with activity in auditory areas) now concerned primarily areas in the superior temporal sulcus, rather than the primary auditory cortex in the superior temporal gyrus (cf. Table 4, with clusters including only a small part of area TE.3).

The pattern of activation associated with the auditory features was also similar between datasets (see Figs. 3A-C rightmost columns and Table 2). As in Exp1, also Exp2a and Exp2b highlighted main foci of activation in the superior temporal gyrus, including most of the primary auditory cortex (see Table 4). Again, the auditory features contrast revealed some modulation of activity in occipital cortex, with significant clusters in medial regions in Exp2a and lateral plus inferior occipital gyrus in Exp 2b. At a lower threshold, Exp2a also showed clusters in the lateral occipital gyrus bilaterally (see Table 2, in italics), identifying the lateral occipital gyrus as the region with the most consistent crossmodal effect of audition in visual occipital cortex.

In summary, feature-based GLMs derived from the computational analyses of the complex and dynamic audio-visual stimuli consistently identified sensory-related responses in occipital and posterior parietal regions for vision and superior temporal gyrus for audition. In addition, in agreement with the condition-based analysis of Exp1, the feature-based GLMs highlighted some cross-talk between modalities; with visual features affecting activity in the superior temporal cortex and auditory features affecting activity in the lateral occipital cortex.

Saliency-based GLM analysis

A main aim of the current study was to identify candidate BOLD correlates of visual and auditory saliency, and – more specifically – to identify any such effect after accounting for changes of low-level sensory features. Thus, within the GLMs that included also the 9 regressors related to visual and auditory features, we tested for the effects of visual and auditory saliency.

For the visual saliency predictor we found consistent effects across the three datasets in the lateral occipital cortex, with activation extending posteriorly to the occipital pole (see Fig. 4, leftmost panels; and Table 3). In Exp 2b, visual saliency was found to co-vary significantly with activity in the right posterior parietal cortex. An analogous cluster was found in Exp2a, but only at a lower statistical threshold (see Fig. 4B central panels, and Table 3). In these areas the BOLD signal increased with increasing saliency, over and above any change induced by features.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

The auditory saliency predictor highlighted a selective and consistent effect in the superior temporal gyrus (see Fig. 4 rightmost column, and Table 3). This included the primary auditory cortex in Heschel’s gyrus (see Table 4), with some overlap between activation related to auditory features and the effect auditory saliency. Nonetheless, in Exps 2a and 2b, which did not include any experimental auditory on/off-sets, the saliency-related effects extended more laterally and anteriorly compared with feature-effect and did not include regions posterior to Heschel’s gyrus (compare Figs. 3B-C versus Figs. 4B-C, rightmost panels). Overall these results indicate that auditory saliency boosted activity in the auditory cortex, over and above any effect of low-level auditory features (i.e. changes of intensity and/or frequency).

Independent component analysis (ICA)

The three datasets (Exp1, 2a, 2b) were submitted to “spatial ICA”, extracting 25 components for Exp1, and 28 components both for Exp2a and for Exp2b. We were able to identify 10 components showing consistent spatial patterns across the three datasets (see Methods section). These are shown in Fig. 5 and in the Supplemental material (Fig. S1). In order to facilitate comparisons between datasets, the ten components were tentatively categorized in two “visual” (V-comp), one “auditory” (A-comp) and seven “others” (O-comp) components.

The first “visual” component included lateral, dorsal and ventral occipital cortex (V-comp1); while the second “visual” component comprised medial occipital regions, including the primary visual cortex (V-comp2). The “auditory” component included superior and middle temporal regions, comprising the primary auditory cortex (A-comp1). The “other” components included primarily high-level associative areas in frontal and parietal cortex (see Fig. S1). Briefly, O-comp1: medial frontal gyrus and bilateral insulae (cf. “salience network”, Seeley et al., 2007); O-comp2: lateral occipital cortex, posterior and intra-parietal cortex, bilaterally; O-comp3: dorsal and ventral fronto-parietal regions, in the left hemisphere (“left memory network”, Damoiseaux et al., 2006); O-comp4: dorsal and ventral fronto-parietal regions, in the right hemisphere (“right memory network”, Damoiseaux et al., 2006); O-comp5: cuneus and precuneus; O-comp6: posterior cingulate, medial frontal cortex and bilateral temporo-parietal junction (“default mode network”, Raichle et al., 2001). O-comp7: ventral occipital cortex and dorsal cerebellum (note: for completeness this component was retained, however it may in fact correspond to an imaging artifact, i.e. cerebrospinal fluid, see Fig. S1).

We sought to highlight possible correspondences between the ICA output and the GLM results. We computed multi-regression r-square values between each ICA component and the GLM regressors associated with visual/auditory features and saliency. The results are displayed in Fig. 6. Concerning features, we found the expected dissociation between the two modalities, with V-comp1 and V-comp2 showing highest r-squares with the visual features (Fig. 6A, first and second rows, columns 1–3), and A-comp1 showing highest r-squares with the auditory features (third row, columns 4–6). These effects were all statistically significant ($p<0.001$, see also legend Fig. 6). These analyses showed also several other significant associations between the ICA components and the GLM features-predictors (see Fig. 6A, values highlighted in bold).

The correlations between the ICA components and the saliency GLM predictors were generally lower (see Fig. 6B). Only in Exp1 did we find a significant relationship between visual saliency and V-comp1 ($p<0.001$; Fig. 6B; first row, column 1). In Exp2a, V-comp1 was again the component with the highest r-squared value, albeit this did not reach full significance at the $p=0.001$ threshold (i.e. $p<0.002$), while in Exp2b the component that correlated most with visual saliency was O-comp2 ($p<0.016$). The pattern of correlations related to auditory saliency was more clear: in all three datasets highest r-square values were

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

associated with the "auditory" ICA component (A-comp1; third row, columns 4–6; all p-values<0.001). This matches the results of the GLM analyses that revealed robust effects of auditory saliency in the superior temporal gyrus (cf. Fig. 4 most right column).

Discussion

We used computationally-derived indexes of visual and auditory bottom–up sensory input (saliency and features) to investigate brain activity during the presentation of complex audio-visual stimuli, and we related this to both a conventional block/event condition-based analysis (cf. Exp 1) and a fully data-driven approach (ICA). Analyses based on visual and auditory features identified changes of BOLD signal in occipital visual cortex and auditory superior temporal cortex, respectively. These patterns were consistent with the results of the condition-based analyses of Exp 1. Visual saliency was found to co-vary positively with activity in extra-striate visual cortex plus the posterior parietal cortex, while auditory saliency was found to boost activity in the superior temporal cortex. The ICA highlighted the implication of several networks during the presentation of the complex audio-visual stimuli. These included "sensory" networks comprising visual and auditory areas, where activity co-varied with the computationally-derived indexes of visual and auditory sensory input. This set of results shows that the combination of computational modeling and GLM enables tracking the impact of bottom–up signals on brain activity during viewing of complex and dynamic multisensory stimuli.

The investigation of brain activity using naturalistic material has been gaining an increasing amount of research interest, with the development of different methodologies for the analysis of these complex datasets (e.g. Hanson et al., 2009; Haxby et al., 2011; Kauppi et al., 2010; Nishimoto et al., 2011). Pioneering work has made use of entirely data-driven methods to highlight critical regions of the brain during vision of complex stimuli. For example, using ICA Bartels and Zeki (2004) identified brain areas that showed distinct fMRI time-courses during movie watching and demonstrated that these temporal patterns are correlated across subjects (see also Hasson et al., 2004). However, purely data-driven methods do not provide us with an explicit way of assessing hypotheses about the role played by brain areas or networks. Hence, effort has focused on new methodologies that seek to identify the specific sensory, motor and/or cognitive aspects of the complex input driving these consistent spatial and temporal patterns. These include comparing ICA results during viewing of complex stimuli vs. rest (Bartels and Zeki, 2005) or between different viewing contexts (e.g. Meda et al., 2009). Others have resorted to hypothesis-driven methods, identifying critical events within the complex stimuli (via scene analysis and/or behavioral measures) and using standard GLM to highlight changes of brain activity associated with these relevant epochs or events (e.g. Wagner et al., 1998; Nardo et al., 2011; see also Hasson et al., 2008).

Using explicit subjective judgments, Bartels and Zeki (2004) investigated attribute selectivity during natural viewing of a movie. Subjects were asked to rate the intensity of perception of color, faces, language and bodies. Ratings were then used as regressors for fMRI analyses (GLM) showing that activity during movie watching correlated with the subjective perception in the expected brain regions: V4 for color, fusiform face area for faces, Wernicke's and auditory areas for speech and the extra-striate body area for bodies. Nonetheless, in order to obtain the subjective ratings, the protocol required interrupting the movie every 2.5–3 min; and – most critically – a single "rating value" had to be associated with an entire 2.5–3 min block of fMRI data (see also Rao et al., 2007, who used motion-related judgments every 6 s, but their ratings were made by a separate group of subjects outside the scanner).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Here we used a more formal approach to quantify the strength of specific sensory features (i.e. not based on subjective ratings or observer-based scene analysis) that also allowed us to track feature-changes on a much shorter time scale (i.e. visual features were initially extracted for each video-frame). For the visual modality, the combination of regressors based on intensity, color, orientation, motion and flicker discontinuities highlighted co-variation with activity in striate and extra-striate occipital areas, plus posterior parietal cortex and inferior temporal areas. Because of the high correlations between predictors we could not identify the specific contribution of each regressor/feature, which is a limitation of the methods presented here (see also Methods section). Nonetheless, comparing the results of the feature-based analysis with a standard block/event approach (cf. Exp 1) indicates that the feature-based analysis suitably identified relevant regions in occipital and posterior parietal cortex. Moreover, the feature-based approach revealed a more restricted (and possibly more selective) set of brain areas compared with the standard block/event analysis, which was found to activate also extensive regions in frontal cortex when comparing motion vs. static conditions.

Bartels et al. (2008) also extracted movie statistics on a frame-by-frame basis for GLM analyses (see also Whittingstall et al., 2010) and considered specifically the effect of local vs. global motion. This enabled identification of motion-related responses in V5/MT (local motion) and, dorsally, in the medial posterior parietal cortex (global motion). Our standard block/event motion-related contrast in Exp1 and all feature-based analyses (Exp1, 2a and 2b) revealed consistent activation both in the lateral occipital cortex (possibly corresponding to V5/MT) and in the posterior parietal cortex. The feature-based analyses included also the ventral occipital cortex, where the block/event-analysis found a small color-selective cluster (possibly corresponding to area V4; see Goddard et al., 2011, who compared blocks of color vs. black-and-white videos), and the striate visual cortex that did not respond to motion in Bartels et al. (2008) but was shown to be strongly modulated by visual contrast in Whittingstall et al. (2010).

Our main contribution here is that we used a biologically-plausible computational model to extract multiple visual features (and auditory features, see below) and concurrently indexed attention-related visual saliency (Itti and Koch, 2000). Hence, we mapped how physical changes of the sensory input affect brain activity and, at the same time, we highlighted the impact of these changes on “higher-level” attentional selection operations. This revealed that visual saliency affected activity in visual cortex and posterior parietal cortex over and above any effect of simple features. Previous studies have associated both the parietal cortex and the occipital visual cortex with the neural representation of visual saliency (Constantinidis and Steinmetz, 2005; Gottlieb et al., 1998; Mazer and Gallant, 2003; see also Nardo et al., 2011; see also Bogler et al., 2011, for related findings using fMRI). In the current study, we demonstrated an effect of visual saliency even after accounting for the influence of low-level features. Hence our current results support the view that saliency-related activity in extra-striate cortex and PPC can be attributed to bottom–up attentional operations beyond mere sensory processing (cf. Bogler et al., 2011; Nardo et al., 2011).

Together with the modeling of visual features and visual saliency, here, for the first time, we used the same computational approach to investigate the representation of auditory features and auditory saliency in the human brain. Behavioral experiments have shown that auditory saliency affects discrimination of both linguistic (Kalinli and Narayanan, 2008) and non-linguistic (Kayser et al., 2005) natural sounds. Previous imaging studies manipulated sounds along several dimensions likely to affect auditory saliency (e.g. Barker et al., 2011; Strainer et al., 1997; Westerhausen et al., 2010), but never formally quantified features and saliency changes at the same time. For example, Strainer et al. (1997) found that increasing the complexity of the auditory input (pure tones, stepped-tones, speech) leads to progressively

Europe PMC Funders Author Manuscripts

greater activation within the primary auditory cortex and to the extension of activation to neighboring auditory association areas. Our fMRI analyses here made use of regressors based on contrasts along different auditory-dimensions: intensity, time, frequency and orientation, with the latter considering concurrent changes in time and frequency (Kalinli and Narayanan, 2007). As for the visual features, due to the high correlation between the feature predictors, we assessed the combined influence of all features together,. This revealed that, overall, these low-level changes affected activity in primary auditory cortex and surrounding areas in the superior temporal cortex. Our results are consistent with recent studies that have also used computational methods to characterize low-level changes in complex sounds and reported co-variations with BOLD activity in the superior temporal cortex (Leaver and Rauschecker, 2010; Lewis et al., 2012). Here we go further by delineating the additional effect of auditory saliency, over and above mere feature-related changes. Compared with the effect of features, the saliency-related modulation extended from Heschl's gyrus anterior-laterally within the superior temporal gyrus. This is in agreement with the spatial distribution of physically-driven responses (here, features) vs. attentionally-driven responses (here, saliency) in the auditory cortex, that previous studies reported in the context of endogenous rather than stimulus-driven attention (Petkov et al., 2004; see also Woods et al., 2009). These findings also fit with the proposal that regions anterior to the primary auditory cortex extract salient object-related information during "scene analysis" of complex auditory input (Lewis et al., 2012).

Together with the effect of visual features/saliency in visual and parietal cortex and the effect of auditory features/saliency in superior temporal regions, we also found some "cross-talk" between the two modalities (see Fig. 3). Crossmodal effects in "sensory-specific" areas have been reported in a variety of contexts (e.g., spatial attention, McDonald et al., 2000; object recognition, Amedi et al., 2005; speech perception, Calvert et al., 2000; see also Brosch et al., 2005 for related studies in primates) and an extensive discussion of these influences is not within the scope of the current study (e.g. see Driver and Noesselt, 2008; Ghazanfar and Schroeder, 2006; Macaluso et al., 2005, for reviews). Nonetheless, it is worth briefly considering the dissociation between de-activation of visual areas by sound versus activation of auditory cortex by vision (cf. standard block/event-analyses in Exp 1). De-activation of visual cortex during unimodal auditory stimulation has been reported before, but this was accompanied by corresponding de-activation in auditory areas during visual stimulation (Laurienti et al., 2002). By contrast here we found that vision enhanced activity in auditory cortex (see also Tanabe et al., 2005). Moreover, Exp2a and Exp2b revealed crossmodal influences of vision in auditory cortex associated with low-level changes of visual features rather attention-related visual saliency. This speaks against a general attentional account (e.g. Macaluso and Driver, 2005), rather suggesting that the influence of vision on audition here may relate to the use of visual input to extract information about the temporal dynamics of the complex stimuli (van Atteveldt et al., 2007), even when the stimuli were presented only visually in Exp 1 (e.g. see Calvert et al., 1997).

In the current study, we used features and saliency-based predictors also to assess components derived from ICA analyses. Fully data-driven ICA revealed 10 components that could be consistently identified across the three datasets. Examination of the spatial patterns lead us to categorize these into two "visual" components, one "auditory" component and 7 "other" components, the latter including associative regions in frontal, parietal and temporal cortices (see Results section, and Fig. S1). Many of these components can be observed also during resting-state fMRI, suggesting that correlated spatio-temporal patterns may arise because of the intrinsic connectivity between regions belonging to each component (Damoiseaux et al., 2006; see also Bartels and Zeki, 2005). Therefore, the issue arises about how to relate ICA patterns with specific functions associated with the processing of the complex stimuli (e.g. Lahnakoski et al., 2012; see also Calhoun and Pearlson, 2012).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Previous studies that have resorted to different combinations of ICA and GLM analyses have, unlike in the current study, typically required presenting stimuli in non-naturalistic conditions (Naumer et al., 2011), involved experimental manipulation of the stimuli (Malinen and Hari, 2011) or relied on subjects' behavior (Meda et al., 2009).

In a previous study, Lahnakoski et al. (2012) investigated the relationship between audio-visual signals and ICA components. Using a combination of manual and automatic stimulus annotations, the authors extracted both low-level features (e.g. auditory entropy) and higher-level categories (e.g. speech vs. music) from cinematographic stimuli. For audition, the results revealed that speech was the main determinant of activity in the two components. One of these, comprising the STG bilaterally, was also sensitive to low-level auditory features, thus exhibiting anatomical and functional characteristics similar to the "auditory" component reported here. For vision, the results highlighted selectivity for low-level features in a component comprising early visual areas, while higher-order stimulus characteristics (head vs. hand vs. mechanical motion) affected other components that included extra-striate areas and associative fronto-parietal cortices.

Here, we used saliency- and feature-regressors to link ICA components with specific aspects of the complex stimuli. This showed that the components tentatively labeled as "visual" indeed contained information about dynamic changes of visual features, while the temporal profile of the "auditory" component correlated with the auditory features. In agreement with the GLM analyses, the same "sensory" components also showed correlation with corresponding saliency predictors. Accordingly, the computationally-derived indexes of visual and auditory bottom–up input enabled us to establish some relationship between ICA components and sensory processes, without making any use of operator-dependent annotations. However, unlike the GLM analyses, the combination of these indexes and ICA did not allow us to highlight the specific contribution of saliency over and above the effects of features.

In summary, we have demonstrated that activity in auditory cortex co-varies with dynamic changes of both auditory low-level features and auditory saliency. In vision, feature-related effects and saliency were found to affect activity in occipital visual cortex as well as in the posterior parietal cortex. By using features and saliency within the same multiple regression model, we demonstrate the contribution of saliency over and above any changes of low-level features, both in visual and in auditory areas. We conclude that the combination of fMRI and computational modeling enables the tracking of visual and auditory stimulus-driven processes associated with the viewing of complex, ecologically-valid multisensory stimuli.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC grant agreement 242809. The Neuroimaging Laboratory, Santa Lucia Foundation, is supported by The Italian Ministry of Health.

References

Altmann CF, Henning M, Döring MK, Kaiser J. Effects of feature-selective attention on auditory pattern and location processing. Neuroimage. 2008; 41:69–79. [PubMed: 18378168]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Amedi A, Von Kriegstein K, Van Atteveldt N, Beauchamp MS, Naumer MJ. Functional imaging of human crossmodal identification and object recognition. Exp. Brain Res. 2005; 166:559–571. [PubMed: 16028028]

Andrade A, Paradis AL, Rouquette S, Poline JB. Ambiguous Results in Functional Neuroimaging Data Analysis Due to Covariate Correlation. 1999

Barker D, Plack CJ, Hall DA. Human auditory cortical responses to pitch and to pitch strength. Neuroreport. 2011; 22:419–433. [PubMed: 21546858]

Bartels A, Zeki S. The chronoarchitecture of the human brain — natural viewing conditions reveal a time-based anatomy of the brain. Neuroimage. 2004; 22:419–433. [PubMed: 15110035]

Bartels A, Zeki S. Brain dynamics during natural viewing conditions — a new guide for mapping connectivity in vivo. Neuroimage. 2005; 24:339–349. [PubMed: 15627577]

Bartels A, Zeki S, Logothetis NK. Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. Cereb. Cortex. 2008; 18:705–717. [PubMed: 17615246]

Bell A, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. Neural Comput. 1995; 7:1129–1159. [PubMed: 7584893]

Bogler C, Bode S, Haynes J-D. Decoding successive computational stages of saliency processing. Curr. Biol. 2011; 21:1667–1671. [PubMed: 21962709]

Brosch M, Selezneva E, Scheich H. Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. J. Neurosci. 2005; 25:6797–6806. [PubMed: 16033889]

Calhoun VD, Pearlson GD. A selective review of simulated driving studies: combining naturalistic and hybrid paradigms, analysis approaches, and future directions. Neuroimage. 2012; 59:25–35. [PubMed: 21718791]

Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. Hum. Brain Mapp. 2001a; 14:140–151. [PubMed: 11559959]

Calhoun VD, Pearlson GD, Pekar JJ. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Hum. Brain Mapp. 2001b; 13:43–53. [PubMed: 11284046]

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iverson SD, David AS. Activation of auditory cortex during silent lipreading. Science. 1997; 276:593–596. [PubMed: 9110978]

Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. 2000; 10:649–657. [PubMed: 10837246]

Constantinidis C, Steinmetz MA. Posterior parietal cortex automatically encodes the location of salient stimuli. J. Neurosci. 2005; 25:233–238. [PubMed: 15634786]

Damoiseaux JS, Rombouts S, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF. Consistent resting-state networks across healthy subjects. Proc. Natl. Acad. Sci. 2006; 103:13848–13853. [PubMed: 16945915]

Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. Neuron. 2008; 57:11–23. [PubMed: 18184561]

Friston KJ. Modes or models: a critique on independent component analysis for fMRI. Trends Cogn. Sci. 1998; 2:373–375. [PubMed: 21227247]

Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage. 2003; 19:1273–1302. [PubMed: 12948688]

Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory? Trends Cogn. Sci. 2006; 10:278–285. [PubMed: 16713325]

Goddard E, Mannion DJ, McDonald JS, Solomon SG, Clifford CWG. Color responsiveness argues against a dorsal component of human V4. J. Vis. 2011; 11

Gottlieb J. From thought to action: the parietal cortex as a bridge between perception, action, and cognition. Neuron. 2007; 53:9–16. [PubMed: 17196526]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Gottlieb JP, Kusunoki M, Goldberg ME. The representation of visual salience in monkey parietal cortex. Nature. 1998; 391:481–484. [PubMed: 9461214]

Greenspan, H.; Belongie, S.; Goodman, R.; Perona, P.; Rakshit, S.; Anderson, CH. Overcomplete steerable pyramid filters and rotation invariance; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 1994; p. 222-228.

Hanson S, Gagliardi A, Hanson C. Solving the brain synchrony eigenvalue problem: conservation of temporal dynamics (fMRI) over subjects doing the same task. J. Comput. Neurosci. 2009; 27:103–114. [PubMed: 19104925]

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. Intersubject synchronization of cortical activity during natural vision. Science. 2004; 303:1634–1640. [PubMed: 15016991]

Hasson U, Furman O, Clark D, Dudai Y, Davachi L. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron. 2008; 57:452–462. [PubMed: 18255037]

Hasson U, Malach R, Heeger D. Reliability of cortical activity during natural stimulation. Trends Cogn. Sci. 2010; 14:40–48. [PubMed: 20004608]

Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge PJ. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron. 2011; 72:404–416. [PubMed: 22017997]

Itti L. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. Vis. Cogn. 2005; 12:1093–1123.

Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res. 2000; 40:1489–1506. [PubMed: 10788654]

Itti L, Koch C. Computational modelling of visual attention. Nature reviews. Neuroscience. 2001; 2:194–203.

Itti, L.; Pighin, F. Realistic avatar eye and head animation using a neurobiological model of visual attention; Proceedings of SPIE 48th Annual International Symposium on Optical Science and Technology; 2003; p. 64-78.

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

Kalinli, O.; Narayanan, SS. A saliency-based auditory attention model with applications to unsupervised prominent syllable detection in speech; Proceedings of InterSpeech; 2007; p. 1941-1944.

Kalinli, O.; Narayanan, S. Combining task-dependent information with auditory attention cues for prominence detection in speech; Proceedings of InterSpeech; 2008; p. 1064-1067.

Kauppi J-P, Jääskeläinen I, Sams M, Tohka J. Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency. Front. Neuroinform. 2010; 4

Kayser C, Petkov C, Lippert M, Logothetis N. Mechanisms for allocating auditory attention: an auditory saliency map. Curr. Biol. 2005; 15:1943–1947. [PubMed: 16271872]

Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol. 1985; 4:219–227. [PubMed: 3836989]

Lahnakoski JM, Salmi J, Jaaskelainen IP, Lampinen J, Glerean E, Tikka P, Sams M. Stimulus-related independent component and voxel-wise analysis of human brain activity during free viewing of a feature film. PLoS One. 2012; 7:e35215. [PubMed: 22496909]

Laurienti PJ, Burdette JH, Wallace MT, Yen YF, Field AS, Stein BE. Deactivation of sensory-specific cortex by cross-modal stimuli. J. Cogn. Neurosci. 2002; 14:420–429. [PubMed: 11970801]

Leaver A, Rauschecker J. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J. Neurosci. 2010; 30:7604–7612. [PubMed: 20519535]

Lerner Y, Honey C, Silbert L, Hasson U. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J. Neurosci. 2011; 31:2906–2915. [PubMed: 21414912]

Lessa PS, Sato JR, Cardoso EF, Neto CG, Valadares AP, Amaro E Jr. Wavelet correlation between subjects: a time-scale data driven analysis for brain mapping using fMRI. J. Neurosci. Methods. 2011; 194:350–357. [PubMed: 20869400]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Lewis JW, Talkington WJ, Tallaksen KC, Frum CA. Auditory object salience: human cortical processing of non-biological action sounds and their acoustic signal attributes. Front. Syst. Neurosci. 2012; 6:1–15. [PubMed: 22291622]

Li Z. A saliency map in primary visual cortex. Trends Cogn. Sci. 2002; 6:9–16. [PubMed: 11849610]

Li Y-O, Adali T, Calhoun VD. Sample dependence correction for order selection in fMRI analysis. IEEE ISBI. 2006:1072–1075.

Liu T, Hospadaruk L, Zhu D, Gardner J. Feature-specific attentional priority signals in human cortex. J. Neurosci. 2011; 31:4484–4495. [PubMed: 21430149]

Macaluso E, Driver J. Multisensory spatial interactions: a window onto functional integration in the human brain. Trends Neurosci. 2005; 28:264–271. [PubMed: 15866201]

Macaluso E, Frith C, Driver J. Multisensory stimulation with or without saccades: fMRI evidence for crossmodal effects on sensory-specific cortices that reflect multisensory location-congruence rather than task-relevance. Neuroimage. 2005; 26:414–425. [PubMed: 15907299]

Malinen S, Hari R. Data-based functional template for sorting independent components of fMRI activity. Neurosci. Res. 2011; 71:369–376. [PubMed: 21925216]

Mazer J, Gallant J. Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. Neuron. 2003; 40:1241–1250. [PubMed: 14687556]

McDonald J, Teder-Salejarvi W, Hillyard S. Involuntary orienting to sound improves visual perception. Nature. 2000; 407:906–908. [PubMed: 11057669]

McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Bell AJ, Sejnowski TJ. Analysis of fMRI data by blind separation into independent spatial components. Hum. Brain Mapp. 1998; 6:160–188. [PubMed: 9673671]

Mechler F, Victor J, Purpura K, Shapley R. Robust temporal coding of contrast by V1 neurons for transient but not for steady-state stimuli. J. Neurosci. 1998; 18:6583–6598. [PubMed: 9698345]

Meda SA, Calhoun VD, Astur RS, Turner BM, Ruopp K, Pearlson GD. Alcohol dose effects on brain circuits during simulated driving: an fMRI study. Hum. Brain Mapp. 2009; 30:1257–1270. [PubMed: 18571794]

Nardo D, Santangelo V, Macaluso E. Stimulus-driven orienting of visuo-spatial attention in complex dynamic environments. Neuron. 2011; 69:1015–1028. [PubMed: 21382559]

Naumer M, van den Bosch J, Wibral M, Kohler A, Singer W, Kaiser J, van de Ven V, Muckli L. Investigating human audio-visual object perception with a combination of hypothesis-generating and hypothesis-testing fMRI analysis tools. Exp. Brain Res. 2011; 213:309–320. [PubMed: 21503649]

Nishimoto S, Vu A, Naselaris T, Benjamini Y, Yu B, Gallant J. Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. 2011; 21:1641–1646. [PubMed: 21945275]

Parkhurst D, Law K, Niebur E. Modeling the role of salience in the allocation of overt visual attention. Vision Res. 2002; 42:107–123. [PubMed: 11804636]

Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL. Attentional modulation of human auditory cortex. Nat. Neurosci. 2004; 7:658–663. [PubMed: 15156150]

Raichle M, MacLeod A, Snyder A, Powers W, Gusnard D, Shulman G. A default mode of brain function. Proc. Natl. Acad. Sci. 2001; 98:676–682. [PubMed: 11209064]

Rao H, Wang J, Tang K, Pan W, Detre JA. Imaging brain activity during natural vision using CASL perfusion fMRI. Hum. Brain Mapp. 2007; 28:593–601. [PubMed: 17034034]

Reichardt W. Evaluation of optical motion information by movement detectors. J. Comp. Physiol. A Sens. Neural Behav. Physiol. 1987; 161:533–547.

Seeley W, Menon V, Schatzberg A, Keller J, Glover G, Kenna H, Reiss A, Greicius M. Dissociable intrinsic connectivity networks for salience processing and executive control. J. Neurosci. 2007; 27:2349–2356. [PubMed: 17329432]

Seghier ML, Price CJ. Dissociating functional brain networks by decoding the between-subject variability. Neuroimage. 2009; 45:349–359. [PubMed: 19150501]

Shamma S. On the role of space and time in auditory processing. Trends Cogn. Sci. 2001; 5:340–348. [PubMed: 11477003]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Strainer JC, Ulmer JL, Yetkin FZ, Haughton VM, Daniels DL, Millen SJ. Functional MR of the primary auditory cortex: an analysis of pure tone activation and tone discrimination. Am. J. Neuroradiol. 1997; 18:601–610. [PubMed: 9127019]

Sui J, Adali T, Yu Q, Chen J, Calhoun VD. A review of multivariate methods for multimodal fusion of brain imaging data. J. Neurosci. Methods. 2012; 204:68–81. [PubMed: 22108139]

Tanabe H, Honda M, Sadato N. Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. J. Neurosci. 2005; 25:6409–6418. [PubMed: 16000632]

Thompson KG, Bichot NP, Sato TR. Frontal eye field activity before visual search errors reveals the integration of bottom–up and top-down salience. J. Neurophysiol. 2005; 93:337–351. [PubMed: 15317836]

van Atteveldt N, Formisano E, Blomert L, Goebel R. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb. Cortex. 2007; 17:962–974. [PubMed: 16751298]

VanRullen R. Visual saliency and spike timing in the ventral visual pathway. J. Physiol. Paris. 2003; 97:365–377. [PubMed: 14766152]

Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science. 1998; 281:1188–1191. [PubMed: 9712582]

Walther D, Koch C. Modeling attention to salient proto-objects. Neural Netw. 2006; 19:1395–1407. [PubMed: 17098563]

Westerhausen R, Moosmann M, Alho K, Belsby S-O, Heamaelaeinen H, Medvedev S, Specht K, Hugdahl K. Identification of attention and cognitive control networks in a parametric auditory fMRI study. Neuropsychologia. 2010; 48:2075–2081. [PubMed: 20363236]

Whittingstall K, Bartels A, Singh V, Kwon S, Logothetis NK. Integration of EEG source imaging and fMRI during continuous viewing of natural movies. Magn. Reson. Imaging. 2010; 28:1135–1142. [PubMed: 20579829]

Woods D, Stecker C, Rinne T, Herron T, Cate A, Yund W, Liao I, Kang X. Functional maps of human auditory cortex: effects of acoustic features and attention. PLoS One. 2009; 4:e5183. [PubMed: 19365552]

Yao H, Shi L, Han F, Gao H, Dan Y. Rapid learning in cortical coding of visual scenes. Nat. Neurosci. 2007; 10:772–778. [PubMed: 17468750]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

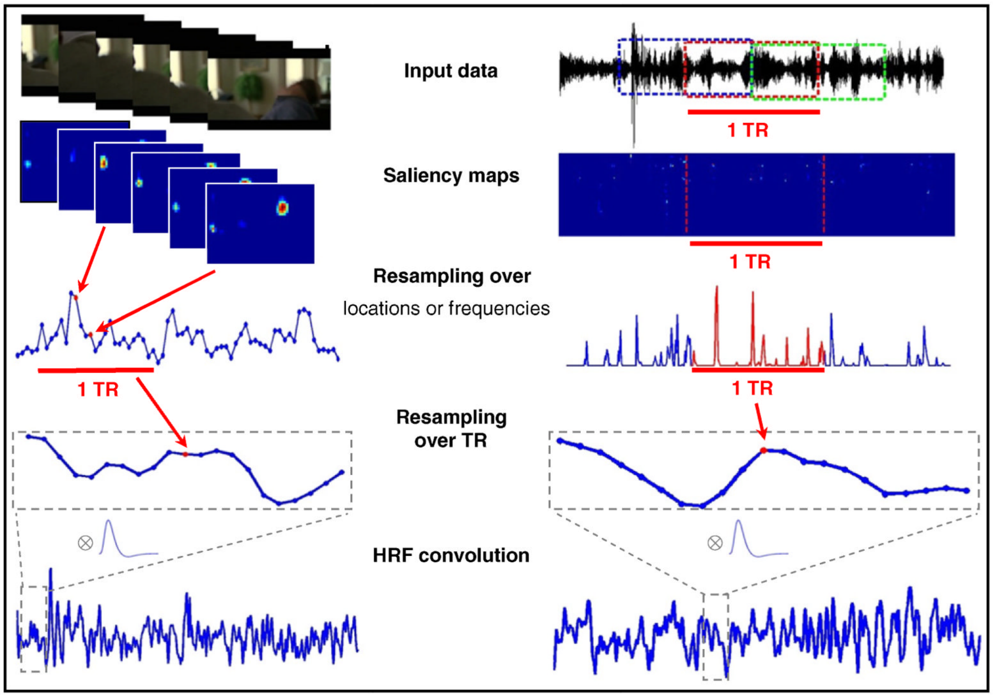


Fig. 1.

Schematic representation of the procedures utilized to construct visual and auditory saliency regressors for the fMRI analyses. Analogous procedures were used to compute the feature-based regressors (see section entitled Parameterization of visual/auditory features and saliency see above for details).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

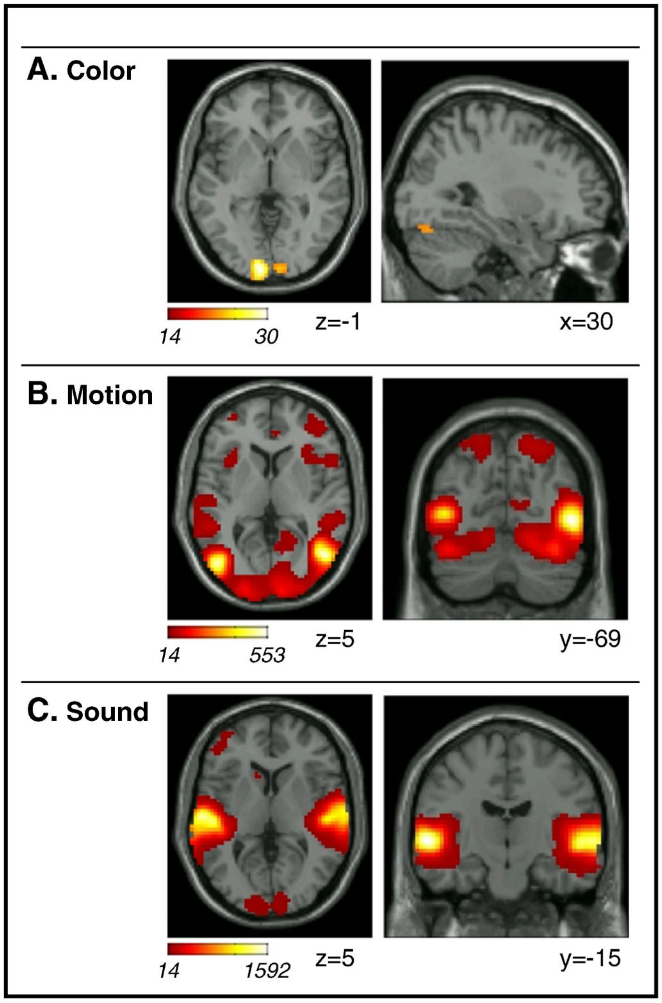


Fig. 2.

Results of the standard block/event GLM analysis of Exp1. Activations associated with F-contrasts testing the combined effect of sustained/block and transient/event onset for: A. Color (color vs. black and white); B. Motion (dynamic vs. static); and C. Sound (sound vs. silence). Activations are projected on transverse and sagittal sections of the SPM8-MNI template. All effects are displayed at a threshold of p-FWE-corr.=0.05 (i.e. with a minimum voxel-level F-value=13.5). Images are in neurological convention (left side of image=left side of brain).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

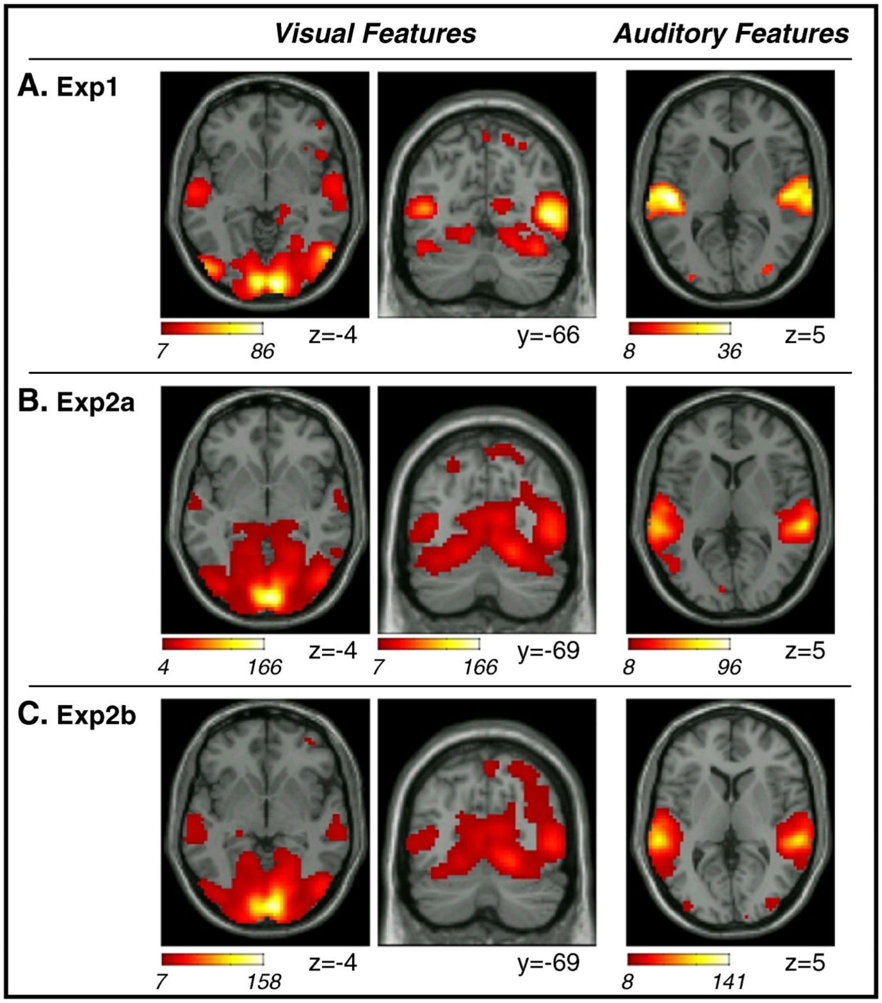


Fig. 3.

Results of the GLM analyses using computationally-derived indexes of dynamic changes of visual and auditory low-level features. Statistical tests considered the combined effects of all features within each modality (F-contrasts). In all three datasets (A. Exp1; B. Exp2a; C. Exp2b), dynamic changes of visual features were associated with activity in striate and extra-striate occipital visual cortex and posterior parietal cortex (cf. coronal sections, center panels). In addition, activity in the superior temporal sulcus also co-varied with visual features (cf. transverse sections, leftmost panels). Auditory features were associated primarily with the superior temporal gyrus (cf. rightmost panels), even though a few voxels were detected also in visual occipital cortex in all three datasets. Activations are displayed at a threshold of p-FWE-corr.=0.05. The corresponding minimum voxel-level F-values were 7.14 (Exp1), 7.01 (Exp2a), 7.02 (Exp2b) for the visual futures; and 8.27 (Exp1), 8.11 (Exp2a), 8.13 (Exp 2b) for the auditory features. Also note that the transverse section of the visual features in Exp2a is displayed at p-unc.=0.001 (minimum F-value=4.11) to show the effect of vision in the superior temporal sulcus (see also Table 2). Images are in neurological convention.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

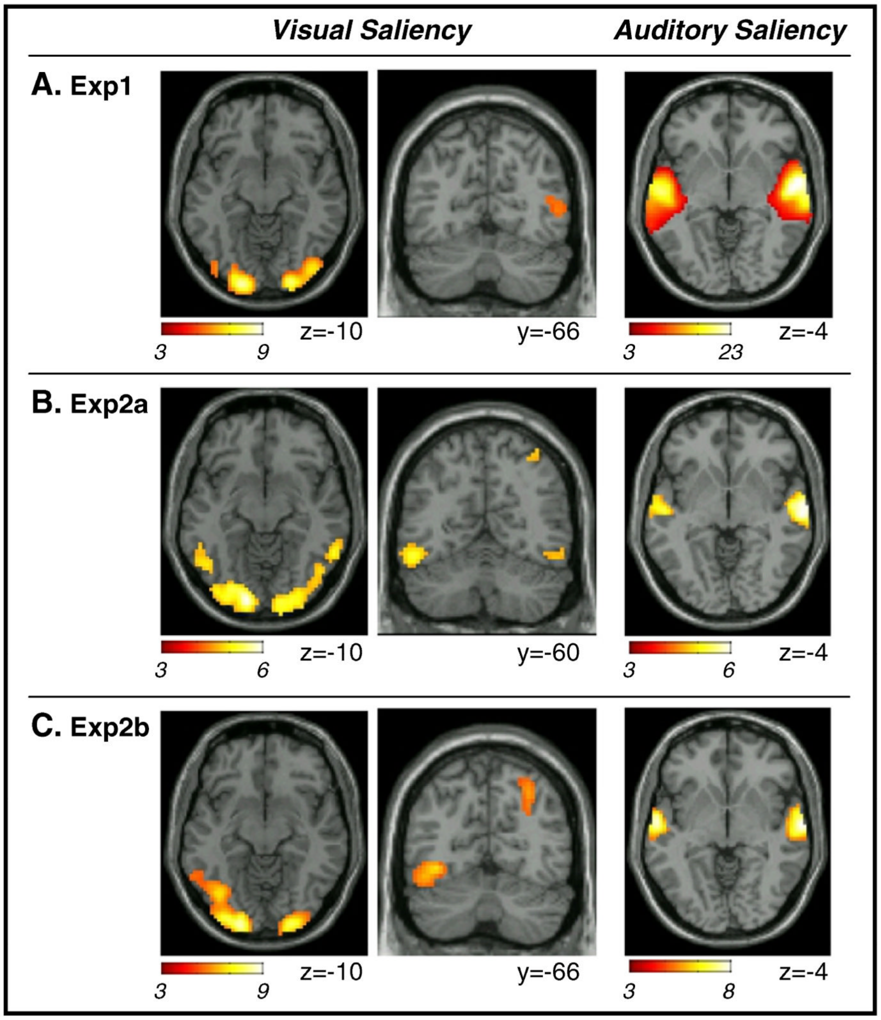


Fig. 4.

Results of the GLM analyses testing for the effects of visual and auditory salience. In all three datasets (A. Exp1; B. Exp2a; C. Exp2b), activity in extra-striate occipital visual cortex co-varied positively with changes of visual salience. In Exp2a and 2b, which made use of the original un-manipulated version of the video, there was an effect of visual salience also in the posterior parietal cortex (see coronal sections in panels B and C). Auditory salience affected activity in the superior temporal gyrus (rightmost panels), where the BOLD signal increased with increasing salience. Activations are displayed at p-FWE-corr.=0.05, cluster-level (minimum voxel-level T-value=3.09, minimum cluster-size=80 voxels), but for the coronal section regarding visual salience in Exp2a (p-unc.=0.001; minimum T-value=3.09, but minimum cluster-size=10 voxels; see also Table 3). Images are in neurological convention.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

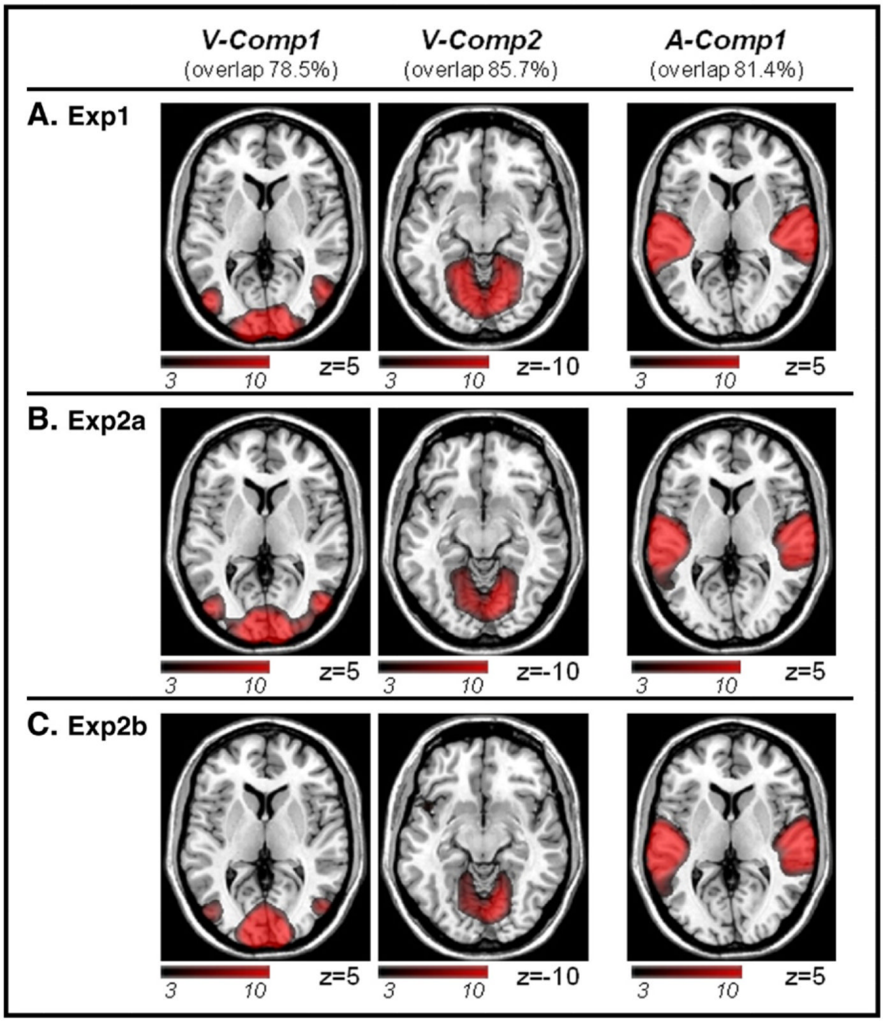
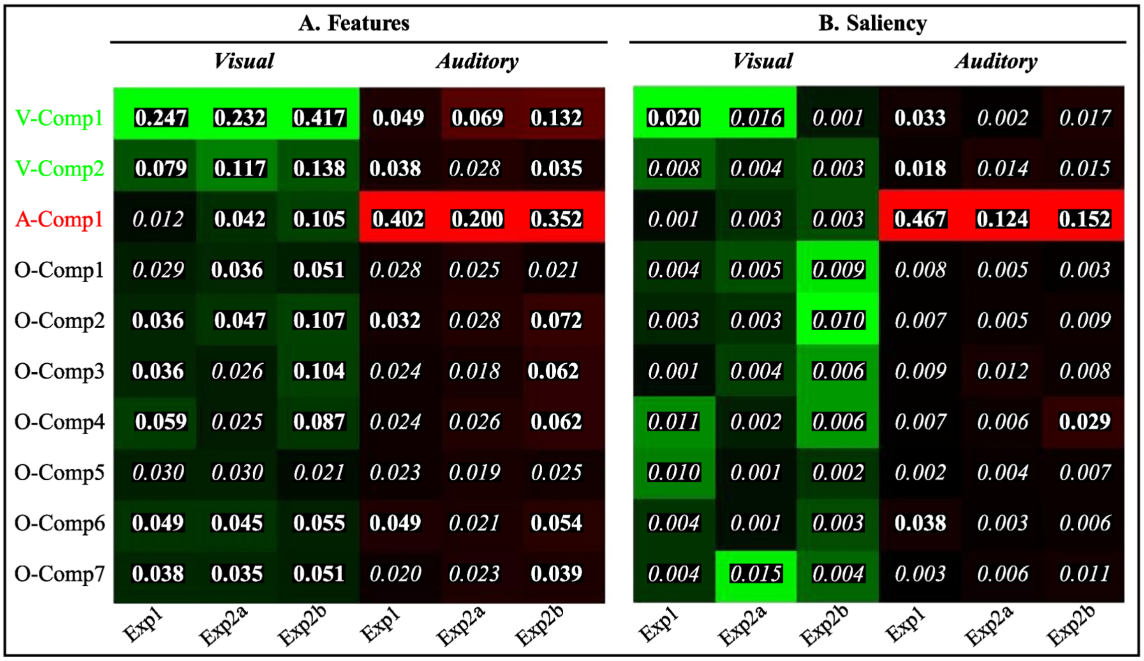


Fig. 5.

The three ICA “sensory” components that were found consistently across datasets (A. Exp1; B. Exp2a; C. Exp2b). For each component, we report a “% overlap” index that quantifies the degree of spatial overlap between the three datasets (see Methods, for details). Two components were labeled as “visual” (V-Comp1 and V-Comp2) and one as “auditory” (A-Comp1), here merely based on the anatomical location of the spatial patterns (please see Fig. 6, for further assessments of the ICA components). Additional ICA components that included higher-order associative regions (O-Comp) are reported in the Supplementary materials. Activations are displayed at p-unc.=0.001. Images are in neurological convention.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



	A. Features						B. Saliency					
	Visual			*Auditory*			*Visual*			*Auditory*		
V-Comp1	**0.247**	**0.232**	**0.417**	**0.049**	**0.069**	**0.132**	**0.020**	*0.016*	*0.001*	**0.033**	*0.002*	*0.017*
V-Comp2	**0.079**	**0.117**	**0.138**	**0.038**	*0.028*	**0.035**	*0.008*	*0.004*	*0.003*	**0.018**	*0.014*	*0.015*
A-Comp1	*0.012*	**0.042**	**0.105**	**0.402**	**0.200**	**0.352**	*0.001*	*0.003*	*0.003*	**0.467**	**0.124**	**0.152**
O-Comp1	*0.029*	**0.036**	**0.051**	*0.028*	*0.025*	*0.021*	*0.004*	*0.005*	*0.009*	*0.008*	*0.005*	*0.003*
O-Comp2	**0.036**	**0.047**	**0.107**	**0.032**	*0.028*	**0.072**	*0.003*	*0.003*	*0.010*	*0.007*	*0.005*	*0.009*
O-Comp3	**0.036**	*0.026*	**0.104**	*0.024*	*0.018*	**0.062**	*0.001*	*0.004*	*0.006*	*0.009*	*0.012*	*0.008*
O-Comp4	**0.059**	*0.025*	**0.087**	*0.024*	*0.026*	**0.062**	*0.011*	*0.002*	*0.006*	*0.007*	*0.006*	**0.029**
O-Comp5	*0.030*	*0.030*	*0.021*	*0.023*	*0.019*	*0.025*	*0.010*	*0.001*	*0.002*	*0.002*	*0.004*	*0.007*
O-Comp6	**0.049**	**0.045**	**0.055**	**0.049**	*0.021*	**0.054**	*0.004*	*0.001*	*0.003*	**0.038**	*0.003*	*0.006*
O-Comp7	**0.038**	**0.035**	**0.051**	*0.020*	*0.023*	**0.039**	*0.004*	*0.015*	*0.004*	*0.003*	*0.006*	*0.011*
	Exp1	Exp2a	Exp2b	Exp1	Exp2a	Exp2b	Exp1	Exp2a	Exp2b	Exp1	Exp2a	Exp2b

Fig. 6.

Graphic summary of the results of the multi-regression analyses between visual/auditory features (A) and saliency (B) and the temporal patterns associated with each ICA component. For each component, feature/saliency index and dataset (Exp1, 2a, 2b) we report the r-square value of the multi-regression. The color-maps are normalized over column to visualize the ICA component that was most related to each index, in each experiment — with brighter colors indicating higher correlations. The r-squared values were transformed into F-values (see Methods) and corresponding p-values were computed. In the figures, r-squared values with p<0.001 are highlighted in bold. The results show that the temporal patterns of “visual” ICA components (V-Comp1 and V-Comp2, see also Fig. 5) co-varied most with visual features (panel A, in green); and that the “auditory” component (A-Comp1) co-varied most with the auditory features (panel A, in red). The “auditory” component also co-varied with auditory salience (panel B, in red), consistent with the results of the GLM analyses showing that both auditory features and salience affected activity in the superior temporal gyrus. Visual saliency was found to co-vary with the temporal profile of V-comp1 (significant r-squared in Exp1; statistical trend in Exp2a, cf. main text).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Areas responding to color, motion or sound in Exp1. F-contrasts jointly tested for sustained/blocked and transient/event effects (p-FWE-corr.<0.05, at the voxel level). Co-ordinates and F-values refer to peak-voxels within each anatomical area. Co-ordinates (mm) are in MNI standard space. Inf/Sup/Post: inferior/superior/posterior.

	H	Coord.			F-val.
Color					
Occipital					
Medial occipital cortex	L	−9	−96	−1	29.98
	R	9	−96	−4	17.54
Sup. occipital cortex	R	12	−99	20	21.64
Inf. occipital cortex	R	30	−75	−19	14.84
Motion					
Occipital					
Medial occipital cortex	L	−9	−99	2	38.45
	R	12	−99	2	121.73
Lateral occipital cortex	L	−48	−75	5	501.91
	R	51	−69	2	552.86
Sup. occipital cortex	L	−18	−96	17	129.10
	R	12	−99	20	149.21
Inf. occipital cortex	L	−33	−78	−19	157.29
	R	33	−78	−19	287.94
Temporal					
Sup. temporal gyrus	L	−66	−21	−1	23.64
Sup. temporal sulcus	L	−51	−45	8	73.30
	R	51	−39	11	77.95
Parahippocampal gyrus	L	−18	−30	−4	21.23
	R	18	−30	−4	35.50
Parietal					
Post. parietal cortex	L	−18	−69	62	33.35
	R	18	−69	59	43.32
Intra parietal sulcus	L	−36	−48	41	24.50
	R	57	−33	41	35.06
Frontal					
Precentral gyrus	L	−51	3	50	5.68
	R	57	−3	47	21.18
Middle frontal gyrus	L	−39	24	38	16.37
	R	39	42	26	21.74
Medial frontal cortex	R	3	30	35	37.31
Insula	L	−42	18	−10	37.76
	R	51	18	−7	39.75
Sound					
Temporal					
Sup. temporal gyrus	L	−60	−15	2	1591.71
	R	60	−9	−4	1513.07
Parahippocampal gyrus	L	−15	−27	−7	22.36

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

	H	Coord.			F-val.
	R	15	−27	−7	26.27
Occipital					
Medial occipital cortex	L	−9	−99	2	32.44
	R	12	−96	17	30.49
Inf. occipital cortex	R	21	−78	−7	20.78
Frontal					
Middle frontal gyrus	L	−39	24	41	21.43
Medial frontal cortex	L	−3	36	35	20.41
Inf. frontal gyrus	L	−42	21	20	20.76
	R	42	21	23	24.87

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 2

Areas where activity co-varied with dynamic changes of visual (top) or auditory (bottom) low-level features. Results are reported for the three independent datasets (p-FWE-corr.<0.05, at the voxel level; unless in italic, then p-unc.<0.001). Co-ordinates and F-values refer to peak-voxels within each anatomical area. Co-ordinates (mm) are in MNI standard space. Inf/Sup/Post: inferior/superior/posterior.

		Exp1				Exp2a				Exp2b			
	H	Coord			F-val	Coord			F-val	Coord			F-val
V-features													
Occipital													
Medial occipital cortex	R	9	−96	−1	85.11	6	−93	−1	166.26	9	−93	−1	158.09
Lateral occipital cortex	L	−48	−75	5	85.92	−45	−78	5	49.29	−48	−78	5	37.62
	R	48	−66	2	79.10	48	−75	5	60.76	48	−72	2	51.78
Sup. occipital cortex	L	−9	−90	41	22.25	−24	−90	32	50.38	−12	−90	35	42.71
	R	18	−90	41	47.88	21	−93	26	79.28	9	−84	41	52.84
Inf. occipital cortex	L	24	−84	−19	44.55	−24	−78	−16	59.15	−18	−51	−7	17.93
	R	33	−81	−19	74.40	15	−81	−13	98.31	21	−60	−10	40.69
Temporal													
Sup. temporal gyrus	L	−54	−18	5	24.36								
	R	50	−18	5	18.25								
Sup. temporal sulcus	L	−60	−18	2	24.77	*− 63*	*− 9*	*− 4*	*7.25*	−60	−24	−4	13.50
	R	60	−12	−1	18.96	*66*	*− 15*	*− 4*	*6.19*	66	−27	−4	14.61
Inf. temporal gyrus	L	−45	−48	−25	24.20	−33	−51	−22	17.75	−60	−60	−16	15.00
	R	42	−54	−25	36.42	57	−54	−13	11.89	57	−54	−13	13.62
Parahippocampal gyrus	L	−18	−30	−7	8.75	−24	−30	−1	7.85	−21	−30	−4	8.75
	R	18	−30	−4	10.86	21	−30	−1	9.87	24	−27	−7	8.96
Parietal													
Post. parietal cortex	R	30	−60	53	8.29	21	−69	59	8.32	30	−69	56	11.10
Intra parietal sulcus	L	*− 36*	*− 48*	*41*	*4.98*	*− 39*	*− 45*	*44*	*4.39*	−39	−45	41	11.88
	R	36	−42	44	9.10	*51*	*− 42*	*47*	*5.14*	51	−45	50	15.21
Inf. parietal cortex	L	−54	−45	8	14.56					−57	−33	26	8.56
	R	57	−39	11	10.57	63	−42	60	16.74	57	−39	20	15.45
Medial parietal cortex	R	3	*− 48*	*47*	*6.15*	6	−51	60	17.61	3	−48	59	15.86
Frontal													
Middle frontal gyrus	L									−45	27	32	12.73
	R	*33*	*36*	*26*	*6.84*					42	33	35	11.82
A-features													
Temporal													
Sup. temporal gyrus	L	−48	−24	5	35.89	−60	−15	−4	92.59	−60	−12	−7	118.29
	R	54	−12	−4	35.39	63	−18	−7	95.68	63	−18	−7	140.78
Occipital													
Medial occipital cortex	L	−3	−81	−7	9.63	−9	−84	5	9.14				
Lateral occipital cortex	L	−33	−90	11	11.83	*− 39*	*− 90*	*5*	*5.73*	−36	−90	11	15.14
	R	30	−78	20	13.14	*27*	*− 87*	*17*	*5.05*	33	−87	20	17.72
Inf. occipital cortex	L	−27	−72	−10	11.25					−27	−48	−7	9.09
	R	*30*	*− 60*	*− 10*	*8.51*					30	−63	−13	18.47
Parietal													
Post. parietal cortex	R	*24*	*− 60*	*53*	*7.18*	*30*	*− 60*	*62*	*5.86*	15	−75	56	13.59
Inf. parietal cortex	R	*66*	*− 30*	*38*	*6.18*					57	−42	41	11.57
Frontal													
Precentral gyrus	L	*− 42*	*0*	*38*	*5.40*	*− 51*	*− 6*	*50*	*7.08*	−45	0	53	14.36
	R	*51*	*9*	*32*	*8.27*	*57*	−6	47	9.28	54	0	47	16.64
Middle frontal gyrus	L					*− 42*	*12*	*23*	*5.97*	−42	15	23	14.20
	R					*36*	*15*	*20*	*5.88*	54	21	29	14.87

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 3

Areas where activity increased with increasing visual or auditory saliency. Results are reported for the three independent datasets (p-FWE-corr.<0.05, at the cluster level; unless in italic, then p-unc.<0.001). Co-ordinates and T-values refer to peak-voxels within the activated clusters. The cluster size (vol, in mm^3) was estimated at a voxel-level threshold of p-unc. =0.001. Co-ordinates (mm) are in MNI standard space. Sup/Post: superior/ posterior.

	H	Exp1					Exp2a					Exp2b				
		Vol		Coord		T-val	Vol		Coord		T-val	Vol		Coord		T-val
V-saliency																
Medial occipital cortex	L	21.7	−15	−99	−4	8.49	34.3	−15	−93	−13	6.03	33.6	−21	−99	−4	8.57
Lateral occipital cortex	L		−42	−87	−4	4.83		−42	−75	−16	3.81		−39	−72	−10	5.33
Medial occipital cortex	R	24.4	24	−99	−4	7.92	25.0	18	−99	−1	6.05	14.2	24	−96	−1	8.44
Lateral occipital cortex	R		42	−84	−10	5.51		45	−78	−19	3.89		39	−90	−10	3.48
Post. parietal cortex	R						*0.6*	*39*	*− 60*	*59*	*3.44*	3.0	30	−66	47	4.27
Precentral gyrus	L	*0.1*	*− 51*	*− 3*	*47*	*3.40*	*1.8*	*− 51*	*18*	*35*	*4.07*	*5.5*	−42	0	32	4.83
A-saliency																
Sup. temporal gyrus	L	84.6	−57	−15	2	20.83	6.4	−54	−15	−1	5.89	9.7	−60	−9	−4	7.28
	R	82.5	60	−6	−4	22.76	8.2	66	−15	−4	6.39	9.8	66	−12	−4	7.70
Middle frontal gyrus	L	3.1	−51	6	44	4.10										

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 4

Extent of activation in auditory areas, in the superior temporal gyrus. Auditory areas are defined according to the SPM Anatomy toolbox. The extents of activation are expressed in percent (%) of each area's volume.

		TE 1.0		**TE 1.1**		**TE 1.2**		**TE 1.3**	
		Left	**Right**	**Left**	**Right**	**Left**	**Right**	**Left**	**Right**
Block-event motion	Exp1	33	–	–	20	–	16	33	17
Block-event sound	Exp1	100	100	100	99	100	100	71	65
Visual features	Exp1	99	97	48	39	77	91	59	59
	Exp2a	–	–	–	–	–	–	5	3
	Exp2b	–	–	–	–	–	–	14	3
Auditory features	Exp1	99	96	97	71	79	91	63	68
	Exp2a	71	61	80	56	90	52	62	65
	Exp2b	91	72	81	65	90	52	63	74
Auditory saliency	Exp1	100	100	100	98	100	107	71	69
	Exp2a	24	8	–	–	37	23	26	33
	Exp2b	18	18	–	–	74	66	33	40