Author Manuscript

Faculty of Biology and Medicine Publication

This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Published in final edited form as:

The context-contingent nature of cross-modal activations of the visual cortex

Pawel J. Matusz[1,2], Chrysa Retsa[1], Micah M. Murray[1,3,4]

[1]The Laboratory for Investigative Neurophysiology (The LINE), Neuropsychology and Neurorehabilitation Service and Department of Radiology, University Hospital Center and University of Lausanne, 1011 Lausanne, Switzerland

[2]Attention, Brain, and Cognitive Development Group, Department of Experimental Psychology, University of Oxford, UK

[3]EEG Brain Mapping Core, Center for Biomedical Imaging (CIBM) of Lausanne and Geneva, 1011 Lausanne, Switzerland

[4]Department of Ophthalmology, University of Lausanne, Jules-Gonin Eye Hospital, Lausanne, Switzerland

Address correspondence to:

Pawel J. Matusz

Neuropsychology and Neurorehabilitation Service,

Centre Hospitalier Universitaire Vaudois (CHUV)

BH7.081, rue du Bugnon 46

1011 Lausanne, Switzerland

Email : pawel.matusz@chuv.ch

Abstract

Real-world environments are nearly always multisensory in nature. Processing in such situations confers perceptual advantages, but its automaticity remains poorly understood. Automaticity has been invoked to explain the activation of visual cortices by laterally-presented sounds. This has been observed even when the sounds were task-irrelevant and spatially uninformative about subsequent targets. An auditory-evoked contralateral occipital positivity (ACOP) at ~250ms post-sound onset has been postulated as the event-related potential (ERP) correlate of this cross-modal effect. However, the spatial dimension of the stimuli was nevertheless relevant in all prior studies where the ACOP was observed. By manipulating the implicit predictability of the location of lateralised sounds in a passive auditory paradigm, we tested the automaticity of cross-modal activations of visual cortices. 128-channel ERP data from healthy participants were analysed within an electrical neuroimaging framework. The timing, topography, and localisation resembled previous characterisations of the ACOP. However, the cross-modal activations of visual cortices by sounds were critically dependent on whether the sound location was (un)predictable. Our results are the first direct evidence that this particular cross-modal process is not (fully) automatic; instead, it is context-contingent. More generally, the present findings provide novel insights into the importance of context-related factors in controlling information processing across the senses, and call for a revision of current models of automaticity in cognitive sciences.

Key words

1. Introduction

The multisensory nature of real-world environments provides obvious benefits for object recognition and goal-directed behaviour. In social situations, with many people speaking, seeing lip movements of the next speaker helps us know where to attend and to understand what will be said next (e.g., van Wassenhove et al., 2005; Zion-Golumbic et al., 2014). Notwithstanding, in laboratory settings even simple sounds are shown to modulate the brain processing and/or facilitate perception of visual objects. At least two prominent types of processes contribute to these effects: multisensory integration of information (reviewed in Ghazanfar and Schroeder, 2006; Stein, 2012; Murray and Wallace, 2012) and orienting of spatial attention to the sound location (McDonald et al., 2000, 2003, 2012; Störmer et al., 2009; reviewed in Koelewijn et al., 2010; Hillyard et al., 2015). Importantly, each of these processes is subject to a differing degree to constraints imposed by the current behavioural goals of the observer, which will determine the efficacy of a particular cross-modal influence. While at least some multisensory processes, such as those based on the detection of multisensory simultaneity, occur independently of the task-relevance of the other-modality signals (Matusz et al., 2011; De Meo et al., 2015; Murray et al., 2015a; Ten Oever et al., in revisions), orienting of involuntary spatial attention might be less impervious to it.

It has been well established within the area of visual attention that even perceptually salient stimuli, if task-irrelevant, fail to attract involuntary shifts of spatial attention (task-set contingent attentional capture; Folk et al., 1992; reviewed in Nobre and Kastner, 2014). This was confirmed by experiments employing brain response measures. Functional magnetic resonance imaging (fMRI) studies have consistently demonstrated that the ventral fronto-parietal brain network that serves as the 'circuit breaker' for the ongoing goal-driven behaviour (i.e., it re-orients attention) responds predominantly, if not exclusively, to 'irrelevant' stimuli as long as these stimuli share features with the target (reviewed in Corbetta and Shulman, 2002). Notably, fMRI evidence has suggested that there are no differences across sensory modalities in engaging the ventral attentional network (in, typically visual, spatial attention tasks; e.g., Downar et al., 2000). However, with their sub-millisecond resolution, event-related potentials (ERPs) might be a method particularly well-suited to study fast-paced, attentional process (e.g., Ding et al., 2014). In line with the behavioural and heamodynamic evidence, ERP studies in visual attention have demonstrated that distracters in spatial attention tasks elicit brain responses indicative of top-down suppression (distracter positivity, Pd), rather than attentional selection (the N2pc component), of those distracters in space (Hickey et al., 2009; Sawaki and Luck, 2010; Wykowska and Schubö, 2010, 2011; McDonald et al. 2012; Gaspar and McDonald, 2014). These findings have jointly suggested that in real-world environments stimuli not matching the current goals of the observer have little ability to attract the observer's attention (with the

exception, maybe, of stimuli whose task-relevance is 'hardwired' in the brain; e.g., Koster et al., 2004; Humphreys and Sui, 2015; Matusz et al., 2015a; Munneke et al., 2015).

Research that employed stimuli from different sensory modalities within visual spatial-attention tasks has been intimating a more nuanced view on this issue. In one exemplary behavioural study, a short sound to the left or right was shown to facilitate perception, as indexed by d', of a faint LED array flash appearing subsequently at the sound location (McDonald et al., 2000). Importantly, a recent pair of studies revealed the likely brain substrates of this cross-modal perceptual benefit. Across a series of experiments, involving both auditory and visual targets, lateralised sounds that preceded these targets were found to elicit positive-going potentials over the contralateral occipital scalp starting at approximately 250ms post-stimulus (ACOP; McDonald et al., 2013; Feng et al., 2014). The positive links between the ACOP amplitude and both subjective and objective measures of perceptual processing, on the one hand, and the fact that the sounds were not predictive (i.e., informative) of target locations, on the other hand, are consistent with shifts of exogenous , involuntary spatial attention underlying the observed cross-modal perceptual benefits (Hillyard et al., 2015).

The task-irrelevance of the ACOP-inducing sounds and the robustness of their effects in perception have opened the possibility that the ACOP, and the exogenous attention orienting it might reflect, is 'automatic' in nature (McDonald et al., 2013). The Miriam-Webster online dictionary defines 'automatic' as a quality: "(…) that allow[s] something to work or happen without being directly controlled by a person". Similarly, a recent review of several models of automaticity as a concept in cognitive research (Moors and de Houwer, 2006) highlights that an automatic process is typically characterised by "features, such as unintentional, uncontrolled/ uncontrollable, goal-independent, autonomous, purely stimulus driven, unconscious, efficient, and fast" (p. 297). Both sources, thus, emphasise predominantly the involuntary nature of an 'automatic' process. The question of automaticity of involuntary shifts of spatial attention is, as we described, hardly new. However, it regains its importance and novelty when considered more broadly, in real-world environments. Here, the multitude of channels providing sensory inputs is mirrored by the multitude of top-down mechanisms that control sensory processing (Doehrmann and Naumer, 2008; Summerfield and Egner, 2009; Nobre and Kastner, 2014). The study of brain and/or cognitive processes at the intersection of these bottom-up and top-down influences, while insurmountable at a first glance, is both feasible and timely; the necessary background has been created by the traditional research involving rigorous experimental setups with unisensory (visual or auditory) stimulation. At the same time, such investigations bring us closer to understanding the information processing as it occurs in situations more closely resembling naturalistic environments.

One notable feature linking all previous empirical reports of the ACOP is that this component has been observed exclusively in response to task-irrelevant sounds that were spatially unpredictable. This opens the possibility that while the ACOP might indeed occur involuntarily, it depends on the stimulus context. The context can be understood as the "immediate situation in which the brain operates" (van Atteveldt et al., 2014) and, more specifically, the observer's expectations. If the circumstances in which the sounds are presented, such as how (un)predictable the sound location is, determine the presence of the ACOP, this would speak against the automaticity of this particular brain/cognitive process. More generally, this would call for a revision of the existing conceptualisations of automaticity of cognitive processes.

While task-relevance is one frequently studied form of top-down control over sensory processing, within (reviewed in Nobre and Kastner, 2014) and across the senses (e.g., Matusz et al., 2011, 2013; reviewed in Talsma et al., 2010; De Meo et al., 2015; Ten Oever et al., in revisions), an increasing number of studies points to similar importance of context-based influences. As demonstrated by traditional, unisensory studies, context influences range from predictions (Summerfield and Egner, 2009), through external and internal states (e.g., remembering something better in a place where one had learnt it), to fine-grained differences in stimulus features (e.g., the object's colour; Bar, 2004; Baddeley et al., 2009). These can affect the activity across scales from a single neuron (reviewed in Gilbert and Li, 2013) to whole-brain cognitive functions, including auditory stimulus parsing, visual search or conditioning (e.g., Saffran et al., 1996; Baker et al., 2004; Courville et al., 2006; Goujon and Fagot, 2013). More recently, the context has been revealed as an important source of top-down control over processing of multisensory information. While some studies demonstrated the role of long-term experience and learning (e.g., Froyen et al., 2009; Stevenson and Wallace, 2013; Barenholtz et al., 2014; Ten Oever et al., 2014; Matusz et al., 2015b), many focused on effects operating at shorter timescales, such as expectations and/or experiences built over the course of a single experimental session (e.g., Murray et al., 2004, 2005; von Kriegstein and Giraud, 2006; Meylan and Murray, 2007; Rosenblum et al., 2007; Beierholm et al., 2009; Powers et al., 2009; Barakat et al., 2013; Chandrasekaran et al., 2009; Thelen et al., 2012, 2014; Matusz et al., 2015c; Altieri et al., 2015), or even across a pair of successive experimental trials (Wylie et al. 2009; Murray et al. 2009; King et al. 2012; Sarmiento et al., 2015). Considered together, the overwhelming evidence for the importance of context-based factors for stimulus processing across the senses and the concomitant limited existing data on the ACOP makes it plausible that irrelevant sounds activate the visual cortex in some contexts but not in others. Verifying the sensitivity of the ACOP to context-based influences defined as expectations was, thus, at the centre of the present study.

More specifically, we investigated whether the ability of irrelevant lateralised sounds to trigger the ACOP depends on the implicit predictability of the location of these sounds. If presence of

the ACOP indeed depends on the unpredictability of the sound location, this would provide strong evidence against the automaticity of these cross-modal activations, as an automatic process would be expected to occur independently of the circumstances. Findings indicative of such sensitivity would likewise have broader implications, in that they would call for consideration and inclusion of top-down control mechanisms based on context in future studies of automaticity of brain and cognitive processes and, more broadly, theoretical models of automaticity within the cognitive sciences. To test our hypothesis, we employed a passive 'oddball' paradigm and measured ERPs elicited by lateralised sounds that were presented while participants watched a muted, subtitled movie. Critically, in some blocks ('spatially irregular contexts') sounds were presented equi-probably to the left versus the right hemispace, while in others ('spatially regular contexts') sounds were located predominantly (80% trials) within one of the two hemispaces (**Figure 1**). The passive setup was employed to further ensure the task-irrelevance of the activation-inducing sounds; in virtually all of the previous reports of the ACOP, the irrelevant sounds that elicited it shared with the targets the lateralised nature of their presentation. This could have rendered the former being perceived as potential targets and thus (rudimentarily) task-relevant. To foreshadow our findings, we have indeed found clear evidence that in our passive paradigm the sounds triggered the ACOP, but exclusively in contexts where their location could not be predicted. These results are a direct demonstration that sounds can activate the visual cortex even when they are not relevant, but these cross-modal activations are dependent on the implicit (un)predictability of the sound location, and due to this 'context-contingency' cannot be regarded strongly automatic.

2. Materials and Methods

2.1. Participants

Fourteen unpaid volunteers provided written, informed consent to participate in the experiment. All procedures were approved by the Ethics Committee of the Faculty of Biology and Medicine of University of Lausanne. Data from 4 subjects were excluded based on poor EEG data quality due to high-amplitude alpha oscillations as well as reported drowsiness during the experiment. The results reported here are from the remaining ten participants (5 women), aged 19–31 years (age mean±SD = 26±4 years). Eight of these subjects were right-handed, and the other two were left-handed (Oldfield, 1971). Because the ERP analyses focused on contralateral versus ipsilateral differences in brain responses (subsequent to collapsing ERPs elicited by left-sided and right-sided stimuli), there should be no influence of participants' handedness. None of the subjects had current or prior neurological or psychiatric illnesses. All had normal or corrected-to-normal vision and reported normal hearing. Some of these data were reported as part of a study focusing on parallel pathways in the auditory

system (De Santis et al., 2007), where no analyses as a function of the spatial location of the sound stimuli were performed.

2.2. Stimuli and procedure

Each subject heard a total of 4 blocks of trials (**Table 1**); the order of which was counter-balanced across participants. As is detailed below, there were four auditory stimuli (2 centre frequencies x 2 perceived positions), whose relative frequency of presentation differed within a block of trials (detailed below). Subjects watched a muted and subtitled film during the experiment, and received no instructions about the auditory stimuli. The employed design engendered implicit expectations in the participants that were based on spatial predictability of sounds in the case of 'spatially regular context' blocks but not (or at least less so) in the case of 'spatially irregular context' blocks. We use the term 'implicit' because there were no explicit demands for the participants to be creating predictions about the perceived spatial location of the sounds.

For the two 'spatially irregular context' blocks, 80% of trials contained sounds of one pitch, irrespective of their perceived location in left or right hemispace, while the remaining 20% of trials were of the other pitch, again irrespective of the sounds' perceived location in the left or right hemispace. The two 'spatially irregular context' blocks fully counter-balanced the preponderant pitch. For the 'spatially regular context' blocks, sounds were presented on 80% of trials at one perceived location, irrespective of their pitch, while the remaining 20% of trials involved sounds presented at the other perceived location, again irrespectively of their pitch. The two 'spatially regular context' blocks fully counter-balanced, in turn, which perceived location was preponderant. That is, the sounds in these blocks were presented equi-probably to the right and the left side. Each block lasted approximately 15 minutes and contained 800 trials.

Auditory stimuli were band-pass filtered noise bursts (100ms duration; 10ms rise/fall; 44100Hz sampling). One stimulus had a 250Hz centre frequency ±¼ octave and the other a 500Hz centre frequency ±¼ octave. The perceived location within the left or the right hemispace was induced by an interaural time difference of 800μs, which led to a perceived lateralisation approximately 90° from central midline. Because these sounds were perceived as emanating from within the listener's head, they provided an important additional test of whether sound-induced activations of contralateral visual cortices are dependent on externally localised sound sources. Stimulus intensity at the ear was approximately 76dB SPL (measured using a CESVA SC-160 sound pressure meter; www.cesva.com). Stimuli were delivered via insert earphones (Etymotic model ER-4P; www.etymotic.com) with a pseudo-randomised inter-stimulus interval of 700-1100ms at steps of 100ms, which was controlled using E-prime (www.pstnet.com/eprime).

Figure 1 and Table 1 here

2.3. Electrophysiological recording and pre-processing

Continuous EEG was acquired at 512Hz through a 128-channel Biosemi ActiveTwo AD-box (www.biosemi.com) referenced to the common mode sense (CMS; active electrode) and grounded to the driven right leg (DRL; passive electrode), which functions as a feedback loop driving the average potential across the electrode montage to the amplifier zero. Data pre-processing, ERP analyses as well as source estimations and analyses were implemented using the CARTOOL software (https://sites.google.com/site/fbmlab/cartool; Brunet et al., 2011). First, the EEG was filtered (high-pass 0.1Hz and low-pass 40.0Hz, respectively, using a second order Butterworth filter with -12db/octave roll-off that was computed linearly in both forward and backward directions to eliminate phase shifts), and segmented into peri-stimulus epochs spanning -100ms pre-stimulus to 500ms post-stimulus onset. Epochs were rejected based on an automated ±80µV artefact rejection criterion as well as visual inspection for eye blinks or other noise transients. For each subject, four auditory evoked potentials (AEPs) were calculated, which we refer to as 'spatially irregular context - left', 'spatially irregular context -right', 'spatially regular context -left' and 'spatially regular context - right'. For example, the spatially irregular context -left AEP included trials from both the 'spatially irregular context' block wherein the 250Hz sound was preponderant (i.e., 80% overall, but only half of these were presented to the left hemispace) and the 'spatially irregular context' block wherein the 500Hz sound was preponderant. The spatially regular context -left AEP included trials from the 'spatially regular context' block wherein left-sided sounds were preponderant, irrespective of their pitch (i.e., 80% of trials).

Data at electrodes with artefacts were interpolated for each subject separately using 3-D splines (Perrin et al., 1987). On average, 12 channels were interpolated (range 5–16 channels). Subsequently, data from the 'spatially irregular context -left' and 'spatially regular context -left' AEPs were re-labelled so that electrodes over the left hemiscalp were treated as if they were located over the right hemiscalp. In this way, data were always coded in terms of their contralaterality (contralateral vs. ipsilateral to the perceived sound location). This likewise allowed for collapsing single-trial data from the left and the right condition, and, hence, to assess the effectiveness of contralateral occipital brain activity to acoustically identical sounds as a function of implicit spatial predictability. Henceforth, we refer exclusively to 'spatially irregular context' and 'spatially regular context' conditions as well as contralateral and ipsilateral scalp sites with respect to the stimuli. The average number (±SEM) of accepted EEG sweeps was 1037±61 for the 'spatially irregular context' condition and 987±59 for the 'spatially regular context' condition. These values did not statistically differ ($t_{(10)}$ = 1.15; p>0.28) and are widely considered as more than sufficient for high signal quality in

ERPs, including lateralised components (Luck, 2005). Prior to group averaging, data were baseline corrected using the 100ms pre-stimulus period, 25 Hz low-pass filtered, and recalculated against the average reference.

2.4. ERP analyses

The ERP analyses followed closely the procedures employed in prior studies on the ACOP (McDonald et al., 2013; Feng et al., 2014). Differences between contralateral and ipsilateral spatially regular and irregular auditory processing were analysed using mean voltages from ERPs over 6 selected parieto-occipital scalp locations from each hemiscalp (see inset in Figure 2; c.f., McDonald et al., 2013; Feng et al., 2014). To allow for the detection of possible changes in the latency of the ACOP in the current study, these ERPs were analysed as a function of time starting from 50ms post-stimulus and using a 2x2 repeated-measures analysis of variance (ANOVA) with within-subject factors of condition (spatially irregular context, spatially regular context) and contralaterality (contralateral, ipsilateral). Only effects persisting for at least 10 contiguous time samples were considered reliable (Guthrie and Buchwald, 1991). While responses from the frequent trials were the focus of the ERP analyses, to further investigate the role of (un)predictability of sound location on the ability of the lateralised sounds to elicit the ACOP, we likewise analysed responses from the rare trials.

2.5. Source estimations

We utilised the local auto-regressive average (LAURA; Grave de Peralta et al., 2001, 2004) distributed linear inverse solution to estimate and statistically compare the likely underlying sources of the effects identified in the above voltage measurements. LAURA is a method for selecting source configurations that mimic the biophysical behaviour of electric vector fields (i.e., activity at one point depends on the activity at neighbouring points according to electromagnetic laws) (details on the forward model and how uniqueness to the inverse problem is achieved are described in Grave de Peralta et al., 2004). The solution space is calculated on a realistic head model that includes 4024 nodes, selected from a 6x6x6mm grid equally distributed within the grey matter of the Montreal Neurological Institute's average brain. It makes no a priori assumptions about the number of sources or their locations and can model multiple, simultaneously active sources. As an output, LAURA provides current density measures; the scalar values of which were evaluated at each node. Given that LAURA is a distributed source model, the issue arises of the possibility of obtaining spurious or 'ghost' sources. A treatment of the validity of LAURA in terms of localisation error is beyond the scope of the present study, though simulations and evaluations of empirical data exist (Michel et al., 2004). We would instead note that determining the mean source estimation across subjects and furthermore statistically comparing these estimations provides one means of minimising the

likelihood of falsely accepting a ghost source as valid, since the probability that a source is consistently observed across individuals and conditions is reasonably small. To correct for multiple comparisons, only nodes with a p-value ≤ 0.05 and clusters of at least 15 contiguous nodes were considered significant. This spatial extent criterion is based on the results of the Alphasim program (available at http:/afni.nimh.nih.gov) and assuming a 4mm FWHM Gaussian smoothing kernel in all directions. This criterion indicates that at a node-level threshold of p≤ 0.05 a cluster of 15 nodes occurs with p≤ 0.005 in a random noise field.

To estimate and statistically assess the sources of the ACOP, contralateral and ipsilateral mean voltages over the 308-330ms post-stimulus period (see Results) were separated and in turn used to generate mirror-symmetric ERPs (see Green et al., 2008 for a similar approach). Voltages at midline electrodes were set to zero. Subsequently, source estimations were compared statistically at each node using a paired t-test. Because the maps were perfectly symmetrical, only sources in the right hemisphere are displayed.

Figure 2. here

3. Results

To investigate the effect of spatial predictability of sound location on sound-induced i – cross-modal activations, AEP voltage values averaged across a set of 6 parieto-occipital electrodes (inset of Figure 2A) were submitted to the 2x2 repeated-measures ANOVA as a function of time. There were significant main effects of condition (412–440ms) and contralatelaterality (478–500ms). Most importantly, there was a two-way condition x contralaterality interaction over the 308–330ms post-stimulus time window. Planned comparisons revealed that the observed interaction was driven by the presence of the ACOP exclusively in the spatially irregular context (308–330ms), but not in the spatially regular context. Figure 2A displays for both contexts contralateral and ipsilateral ERPs averaged over the 6 electrodes, and Figure 2B displays for both contexts the scalp topography at the time of the peak of the ACOP in the spatially irregular context (with left and right hemiscalps showing ipsilateral and contralateral activity, respectively). We likewise tested the effects of spatial predictability on the AEPs elicited by the rare sounds in each context. The two main effects were reliable (condition: 88–154ms; contralaterality: 136–166ms; see Supplementary Figure 1). However, there was no evidence for a two-way condition x contralaterality interaction.

Source estimations were calculated for the AEPs from the spatially irregular context over the 308–330ms post-stimulus period (i.e., the time period yielding a significant ACOP as measured at the scalp surface). Estimated source activity was significantly stronger within the contralateral versus

ipsilateral hemisphere. Differential source activity was localised to clusters within the lateral occipital cortex (locally minimum p-value at 29, -74, 22mm) and precuneus (locally minimum p-value at 17, -56, -32mm using the Talairach and Tournoux (1988) atlas) (**Figure 2C**). These loci are in line with previous research (cf., Figure 2c in McDonald et al., 2013; also Feng et al., 2014). Notably, there was no evidence for concomitant differential activity within the auditory cortices.

4. Discussion

Salient sounds can elicit activations of the contralateral occipital cortex ~300ms after sound onset. Such cross-modal activations are observed despite the sounds being uninformative about the location of the subsequent task-relevant stimuli. We provide the first direct evidence that the ACOP is observed in some experimental contexts, but not others. This 'context-contingency' of the ACOP goes against its supposed strong automaticity. In the following sections, we discuss how the present findings advance our understanding of the ACOP and spatial attention orienting more generally. We likewise delineate the broader implications of these results for our understanding of top-down control and automaticity, respectively, and how these advances were made possible by studying the two phenomena in complex, multisensory settings resembling naturalistic environments.

Specifically, our results have shown that whether sounds activated the visual cortex depended on the implicit (un)predictability of their location. Using an analytical approach closely resembling that employed in the previous studies, we too found that lateralised sounds (now completely task-irrelevant because they were presented within a passive paradigm) reliably triggered late-latency activity within contralateral visual cortices (i.e., over the 308-330ms post-stimulus time window). This timing is consistent with the latency of the ACOP observed in prior studies (McDonald et al., 2013; Feng et al., 2014). Likewise, the sources of this activity were within loci similar to those in the previous reports; though we would additionally highlight here that we provide statistical evidence for stronger sources within the contralateral versus ipsilateral hemisphere, whereas the prior studies estimated sources based solely on group-averaged data. This notwithstanding, the critical novel finding of our study is that the ability of irrelevant lateralised sounds to activate the contralateral visual cortex is 'context-contingent', i.e., it depends on the sounds occurring in a particular context. Specifically, in our study the ACOP was observed exclusively in situations where these sounds were largely unpredictable in their location in space. When spatially predictable, the very same sounds failed to elicit the ACOP. Importantly, the analyses performed here enabled us to reveal that, generally speaking, the ACOP reflects an enhanced occipital cortex response contralaterally to the sound location. As clearly visible in Figure 2A, the predictability of the location of the irrelevant sounds (that enabled their inhibition; see below) results in the suppression of the

contralateral activity elicited by these sounds, so that it is resembles the activity triggered by the same sounds across the ipsilateral hemisphere. This pattern of cross-modal enhancements and their top-down suppression (where possible) complement and extend, respectively, the previous evidence from active-task settings (McDonald et al., 2013; Feng et al., 2014) that shifts of exogenous, involuntary spatial attention underlie the ACOP.

4.1. The importance of context-based control for information processing in multi-sensory environments

The present results shed new light on the importance of top-down control related to the context in which stimuli are processed in multisensory environments; in particular, the observer's expectations about the stimuli in the environment and their statistical attributes. The critical manipulation in the present study was the regularity with which sounds appeared in one versus another hemispace. The aim of the manipulation was to induce in the participants predictions regarding the sound location, with these predictions being implicit in nature, as the sounds were presented within a completely passive paradigm. The efficacy of stimulus regularity in altering sensory processing has been repeatedly documented despite anaesthesia, sleep, and task-irrelevance of the stimuli, and reflected by presence of the mismatch-negativity responses (MMN; Escera et al., 2014; Fishman, 2014; Näätänen et al., 2014; Schröger et al., 2014; Todd et al., 2014). However, the current results cannot be interpreted within the MMN framework. There was no evidence of an ACOP in the ERPs elicited by the rare sounds. The responses to standard stimuli were those that clearly demonstrated the context-contingent nature of the cross-modal visual cortex activations. Thus, our findings shed important, novel light onto the role of top-down mechanisms based on context in controlling the efficacy of processing of cross-modal inputs.

The literature is rife with examples of the brain's capacity to improve perception of relevant objects in the environment by utilising the available 'task-irrelevant' information. The idea was popularised by a variety of studies from within the field of visual attention. For example, research on the contextual cueing (e.g., Chun and Jiang, 1998) has been consistently demonstrating how regularities in 'irrelevant' stimulation, such as consistency in the location of distracters within visual-search arrays, are utilised to facilitate search behaviour. Similar benefits have been observed for the foreknowledge regarding a temporal delay between, otherwise uninformative, distracters and subsequent visual-search targets (Watson et al., 2003). However, real-world situations are nearly always both structured *and* multisensory in nature. Thus, what helps us better understand sensory processing within ethologically valid settings are investigations into how context-based top-down mechanisms, such as expectations, control the processing of stimuli that appear across the senses. A growing literature reveals the importance of context-based factors for the efficacy of processing in

multisensory settings. In the case of naturalistic stimuli, e.g., speech, the natural delay between the visual and the subsequent auditory input is well known to facilitate comprehension (e.g., van Wassenhove et al., 2005). Recently, the mechanisms have been revealed by which a cross-talk between higher-level (language and top-down attention areas) areas and lower-level auditory cortices enables, currently unattended but potentially important, stimuli to become attended in real-world like, multi-speaker settings (i.e., the 'cocktail party' context; Zion-Golumbic et al., 2014). The fact that the present effects, obtained with simple acoustic information (tones), were localised to lateral occipital cortices is consistent with these findings and supports the notion of a privileged interplay between auditory cortices, involved in complex sound recognition (and, perhaps, also localisation), and visual cortices, involved in visual-object identification (often person recognition) (e.g., Blank et al., 2014). Studies using simplistic (but carefully controlled) experimental paradigms have been documenting for *de facto* decades the benefits of perception in cross-modal settings (reviewed in, e.g., Spence and Driver, 2004). Similarly, if context is understood as 'the immediate situation' (van Atteveldt et al., 2014) within which a stimulus is processed by the brain, the literature offers a plethora of examples whereby the brain utilises regularities (short-term as well as long-term) within and across the senses to facilitate the processing of the current inputs that are relevant to the task-at-hand (e.g., von Kriegstein and Giraud, 2006; Rosenblum et al., 2007; Beierholm et al., 2009; Powers et al., 2009; Barakat et al., 2013; Chandrasekaran et al., 2009; Thelen et al., 2014; Matusz et al., 2015c; Altieri et al., 2015). For example, the efficiency of the perceptual processing can be increased, as indexed by early-latency reductions in the brain-response strength, after several days of explicit, object-discrimination training with multisensory stimuli (Altieri et al., 2015). Similarly, learning new face-voice pairings benefits from long-term experience (Barenholtz et al., 2014).

Our results fit with these existing findings while at the same time extending them in (at least) one important way. Specifically, when the current findings are considered within the extant literature, they suggest that the current behavioural goals of the observer (i.e., top-down attentional control mechanisms) are what mediate the influence of top-down context-based control mechanisms, such as expectations, on the efficacy of cross-modal stimuli. In the previous reports of the ACOP, the spatially uninformative nature of the distracter sounds rendered them largely task-irrelevant (McDonald et al., 2013; Feng et al., 2014). Nonetheless, the fact that the spatial attribute of their presentations was nearly always shared with the targets suggests that the participants could find the distracter sounds 'rudimentarily' relevant. The positive links between the ACOP amplitudes and the subjective perceptual judgements (McDonald et al., 2013) and the objective improvements in performance on visual targets (Feng et al., 2014) alike could be taken as evidence in support of this hypothesis (cf., the task-contingent nature of activations within the ventral fronto-parietal network argued to be a brain substrate of the exogenous, involuntary attention-orienting system; Corbetta

and Shulman, 2002). The task-free setup of the current study was advantageous inasmuch as it rendered the sounds completely irrelevant and, thus, enabled us to reveal the importance of context-based top-down control (by successfully instilling fore-knowledge about the distracter location) for their ability to elicit the ACOP. Our results are most easily explained by the implicit predictability of the sound location having been utilised by the brain, when available, to inhibit the processing of the completely irrelevant sounds. In turn, in contexts where sounds were presented equally likely to left and right, the same information was simply not available, what in turn prevented the inhibition of the exogenous attention shifts that give rise to the ACOP (Hillyard et al., 2015). The idea that the sound-induced activity has been suppressed when possible is supported by findings from visual attention suggesting that, compared to irregular/unstructured stimuli, regular/structured stimuli are involuntarily assigned greater attentional priority (Zhao et al., 2013). In other words, spatially regular/predictable sounds might have been found particularly distracting by the participants. The role of goals in mediating the effect of expectations on multi-sensory processing was in fact demonstrated recently by Sarmiento et al. (2015) in a study on how judgements of visual stimulus durations are affected by simultaneous albeit irrelevant sounds. The cross-modal influence of sounds was modulated by the visual stimulus location, differing based on whether a particular location was associated with a low versus high incidence of auditory-visual (in)congruence, as well as by the n-1 trial history, i.e., changes in or maintenance of the particular context. Thus, the utilisation of statistical regularities (i.e., context) for goal-directed behaviour can even multiplex across multiple time scales simultaneously.

In summary, considered together with the existing literature, our data reveal that regularity in cross-modal stimulation acts as a 'double-edged sword' regarding the efficacy of the irrelevant stimuli. When helpful to the goal-directed behaviour, regular/predictable stimuli will be continuously processed (i.e., selected) and utilised by the brain (e.g., Ten Oever et al. 2014). However, stimulus regularity likewise seems to enable the observer to effectively inhibit irrelevant stimuli in situations where they are highly unlikely to directly facilitate the ongoing behaviour, e.g., in no-task setups, as in the current experiment. To summarise, our results demonstrate that the consequences of stimulus regularity on cross-modal processing are mediated by the current goals of the observer. This, however, should only hold true as long as the mechanism underlying the processing of these stimuli is dependent on the context. The sounds' predictability, together with the context-contingent nature of the exogenous attention shifts (and of the ACOP), is what made the top-down inhibition of the sounds possible in the present study.

4.2. Investigations of sound processing in complex, multisensory settings

The current results indeed offer insights into the mechanisms underlying the cross-modal visual cortex activations as well as the exogenous attention shifts that likely underlie them. First, one important way in which our results replicate and extend the past findings on the ACOP is that they confirm it occurs independently of the task-relevance of the sounds that induce it. Despite the complete irrelevance of the auditory stimuli ensured by the task-free nature of the paradigm employed here, the cross-modal visual cortex activations were reliably observed, with the latency, topography and the brain sources consistent with previous reports (c.f., McDonald et al., 2013; Feng et al., 2014). Second, and more importantly, the context-contingence of the ACOP revealed by the current results is difficult to reconcile with its 'automatic' or even strongly automatic (see below) nature. In naive terms, an 'automatic' process should be characterised by its ubiquitous nature, i.e., the ACOP should be observed whenever one is presented with sounds that are lateralised (similarly to other, nominally visual ERP components related to attentional processes, such as Pd or N2pc; e.g., Hickey et al., 2009; see also e.g., Matusz and Eimer, 2013) and task-irrelevant (if exogenous, involuntary spatial-attention shifts are the underlying mechanism). In fact, a closer inspection of the results of Feng et al. (2014) would already provide a first indication that the ACOP may not be fully automatic. *A posteriori* sorting of trials according to performance on a visual discrimination task revealed that the ACOP was triggered by sounds on some, but not all, trials; the ACOP was absent on the trials where the sound and the visual target appeared at the same location but the judgments were incorrect. Our results demonstrate that the ACOP can be directly abolished, if one manipulates the context in which the sounds are presented.

More generally, ours and prior studies of the ACOP highlight the importance of establishing a clear criterion, or, more likely, criteria for dubbing a process 'automatic' in nature. Based on the review of several models of automaticity within the cognitive psychology, Moors and de Houwer (2006) proposed that the automaticity of a cognitive (and, by extrapolation, a brain) process can only be relative, rather than absolute. This notion is in accordance with more general findings that even archetypically 'automatic' processes, such as reflexes, are modulated by top-down factors (e.g., Mathôt and Van der Stigchel, 2015). Importantly, Moors and de Houwer (2006) proposed also a set of such criteria. Their first criterion is the occurrence of the process despite the task-irrelevance of the eliciting stimuli. As already discussed, our results provide strong evidence that the ACOP fulfils this criterion, because the sounds we employed were completely irrelevant as there was simply no task. The second proposed criterion, related to the first one, postulates that the presence of the process should be independent of the demands of the current task. Previous studies have observed the ACOP across a variety of tasks, which differed in the sensory modality of the target as well as the difficulty of the task itself (i.e., detection vs. discrimination), suggesting the ACOP might indeed be robust against the task demands. The third proposed criterion is that the process occurs without

conscious awareness of the eliciting stimuli. Additional research will be required to evaluate this in the case of the ACOP.

The fourth criterion concerns the timing of a given process, which is considered a proxy for the presumed hierarchical level in the brain at which the process occurs and, by extension, its susceptibility to top-down control processes. Current evidence regarding the timing of inputs *between* sensory cortices has demonstrated a surprising degree of interplay, wherein auditory-driven responses within nominally visual cortices coincide with and sometimes even precede visually-driven responses to the same external event (Schroeder et al., 2004; Musacchia and Schroeder, 2009; see also Raij et al., 2010, and Brang et al., 2015, for MEG and ECoG evidence in humans, respectively). Consequently, cross-modal effects at the brain level can transpire at latencies still considered within the initial 'sweep' of sensory-driven brain activity that is traditionally considered largely automatic (e.g., Lamme and Roelfsema, 2000). Accordingly, sounds can modulate activity within visual cortices within the first 100ms post-stimulus onset, can directly impact behaviour, and can do so outside the listener's conscious awareness (Romei et al., 2007, 2009; Spierer et al., 2013; reviewed in Murray et al., 2015a). In line with this general notion, the multisensory processing that occurs within the first 100ms post-stimulus is increasingly revealed to do so largely independently (at least in its presence) of the top-down control of goals, semantics, or stimulus context (the eMSI; De Meo et al. 2015; see also Murray et al., 2015a; Ten Oever et al., in revisions). In this regard, the timing of the onset of the ACOP at ~250–300ms is considerably later than the timing of other cross-modal and multisensory processes that could be considered automatic inasmuch as they occur independently of a multitude of top-down influences.

As we have argued, goals are not the only type of top-down control. In respect to the present results, we propose that the unpredictable nature of sound location might be a necessary condition for task-irrelevant sounds to be capable of attracting shifts of reflexive spatial attention and, as a result, activate higher-order visual cortices. In real-world environments, the brain typically utilises information about both the task-relevance of and the regularities in the stimulation in the environment. In line with our argument, when studied in such ethologically valid settings, the ACOP was indeed found to be susceptible to top-down control. Just as the attentional capture by unisensory, visual distracters in spatial visual-attention tasks has been revealed to be 'task-set contingent' (i.e., determined by the observers' goals, Folk et al., 1992), we have demonstrated that the ACOP is 'context-contingent'. On the one hand, such findings support the view that the ACOP reflects shifts of exogenous, 'involuntary' spatial attention. Further evidence is provided by the observation of a positive correlation between the ACOP amplitude and the subjective as well as the objective measures of visual perception (reviewed in Hillyard et al., 2015). On the other hand, our findings bear important theoretical implications for the current understanding of the necessary

criteria for a brain/cognitive process to be considered strongly automatic, particularly if this process is to be understood in real-world environments where different forms of top-down control interact with each other to facilitate purposeful behaviour. Specifically, existing models of automaticity (see Moors and de Houwer, 2006) might be enriched by attention to context-based influences; among which the observer's expectations are but one, notable example. If a process occurs irrespectively of the circumstances or 'the immediate situation' (van Atteveldt et al. 2014) in which the brain operates, the process is more likely to be strongly automatic. Generally, it may be more valid to instead specify the conditions under which a given neural event may be elicited or not elicited, rather than to classify it into semi-binary categories. Furthermore, our results open an exciting possibility that automaticity of brain/cognitive processes and its dependence on experience (Astle and Scerif, 2011; Amso and Scerif, 2015; Murray et al., 2015b) can perhaps be more accurately understood by testing them in settings that resemble more naturalistic environments (e.g., Matusz et al., 2015b). In such environments, there is a variety of dimensions along which the stimuli that are present at any point time differ. Variations in task relevance of the sensory inputs are accompanied by and interact with variations in knowledge about/ experience with those stimuli, with the end-result on sensory processing depending on the nature (unisensory, cross-modal and multisensory) of the inputs themselves.

4.3. Conclusion

To summarise, the present findings critically advance our understanding of the top-down mechanisms that control the ability of irrelevant sounds to influence visual cortex activity. While task relevance has been traditionally perceived as the predominant top-down mechanism determining the efficacy of such cross-modal influences, we reveal that this also applies to statistical stimulus regularities. The extraction of regularities from the presentation of such irrelevant stimuli, likely occurring in an involuntary fashion, will benefit the processing of task relevant stimuli in the case of presence of a task. However, in contexts where the lack of task renders the structured sounds completely irrelevant, their regularity and, thus, implicit predictability, will allow the observer to successfully inhibit them (if the mechanism underlying their cross-modal influences is context-contingent, as in the case of the ACOP). In other words, statistical regularity is a 'double-edged sword' that enables effective suppression of the processing of sounds (i.e., their attentional selection) in situations where irrelevant stimuli are highly unlikely to benefit the current behaviour. More generally, our findings open novel, exciting avenues to understand the nature of specific processes (cross-modal and otherwise) and their nature, by testing them in situations where factors related to task relevance and context of their occurrence are both systematically manipulated.

5. Acknowledgements

6. References

Altieri N, Stevenson RA, Wallace MT, Wenger MJ (2015) Learning to associate auditory and visual stimuli: Behavioral and neural mechanisms. Brain Topogr. 28: 479-93.

Amso D, Scerif G (2015) The attentive brain: Insights from developmental cognitive neuroscience. Nature Rev. Neurosci. 16: 606-619.

Astle DE, Scerif G (2011) Interactions between attention and visual short-term memory (VSTM): What can be learnt from individual and developmental differences? Neuropsychologia 49: 1435-1445.

Baddeley A, Eysenck MW, Anderson, MC (2009) Memory. Hove: Psychology Press.

Baker CI, Olson CR, Behrmann M (2004) Role of attention and perceptual grouping in visual statistical learning. Psyc. Sci. 15: 460-466.

Bar M (2004) Visual objects in context. Nat. Rev. Neurosci. 5: 617–629.

Barakat BK, Seitz AR, Shams L (2013) The effect of statistical learning on internal stimulus representations: regular items are enhanced even when not predicted. Cognition 129: 205–211.

Barenholtz E, Lewkowicz DJ, Davidson M, Mavica L (2014) Categorical congruence facilitates multisensory associative learning. Psyc. Bull. Rev. 21: 1346–1352.

Beierholm UR, Quartz SR, Shams L (2009) Bayesian priors are encoded independently from likelihoods in human multisensory perception. J. Vis. 9: 23.

Blank H, Wieland N, von Kriegstein K (2014) Person recognition and the brain: Merging evidence from patients and healthy individuals. Neurosci. Biobeh. Rev. 47: 717–734.

Brang D, Towle VL, Suzuki S, Hillyard SA, Di Tusa S, Dai Z, ..., Grabowecky, M. (2015). Peripheral sounds rapidly activate visual cortex: Evidence from electrocorticography. J Neurophys. DOI: 10.1152/jn.00728.2015

Brunet D, Murray MM, Michel CM (2011) Spatiotemporal analysis of multichannel EEG: CARTOOL. Comput. Intel. Neurosci. 2012: 2.

Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. PLoS Comput. Biol. 5: e1000436.

Chun MM, Jiang Y (1999) Top–down attentional guidance based on implicit learning of visual covariation. Psyc. Sci. 10: 360–365.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nature Rev. Neurosci. 3: 201-215.

Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. Trends Cogn. Sci. 10: 294-300.

De Meo R, Murray MM, Clarke S, Matusz PJ (2015) Top-down control and early multisensory processes: Chicken vs. egg. Front. Integr. Neurosci. 9:17. doi: 10.3389/fnint.2015.00017.

De Santis L, Clarke S, Murray MM (2007) Automatic and intrinsic auditory "what" and "where" processing in humans revealed by electrical neuroimaging. Cer. Cortex 17: 9–17.

Ding Y, Martinez A, Qu Z, Hillyard SA (2014) Earliest stages of visual cortical processing are not modified by attentional load. Hum. Brain Map. 35: 3008-3024.

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

Downar J, Crawley AP, Mikulis DJ, Davis KD (2000) A multimodal cortical network for the detection of changes in the sensory environment. Nature Neurosci. 3: 277–283.

Escera C, Malmierca MS (2014) The auditory novelty system: An attempt to integrate human and animal research. Psychophysiology 51:111–123.

Feng W, Störmer VS, Martinez A, McDonald JJ, Hillyard SA (2014) Sounds activate visual cortex and improve visual discrimination. J. Neurosci. 34: 9817–24.

Fishman YI (2014). The mechanisms and meaning of the mismatch negativity. Brain Topogr. 27: 500–526.

Folk CL, Remington RW, Johnston JC (1992) Involuntary covert orienting is contingent on attentional control settings. J Exp. Psyc. Hum. Perc. Perf. 18: 1030-1044.

Froyen D, van Atteveldt N, Bonte M, Blomert L (2009) Crossmodal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. Neurosci. Lett. 430: 23-28.

Gaspar JM, McDonald JJ (2014) Suppression of salient objects prevents distraction in visual search. J Neurosci. 34: 5658-5666.

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

Gilbert CD, Li W (2013) Top-down influences on visual processing. Nature Rev. Neurosci. 14: 350-363.

Goujon A, Fagot J (2013) Learning of spatial statistics in nonhuman primates: contextual cueing in baboons (Papio papio). Beh. Brain Res. 247: 101-109.

Grave de Peralta Menendez RG, Andino, SG, Lantz G, Michel CM, Landis T (2001) Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. Brain. Topogr. 14: 131–137.

Grave de Peralta Menendez RG, Murray MM, Michel CM, Martuzzi R, Gonzalez Andino, SL (2004) Electrical neuroimaging based on biophysical constraints. Neuroimage 21 527–539.

Green JJ, Conder JA, McDonald JJ (2008) Lateralized frontal activity elicited by attention-directing visual and auditory cues. Psychophysiology 45: 579–587.

Harms L, Fulham WR, Todd J, Budd TW, Hunter M, Meehan C, ... Michie PT (2014) Mismatch negativity (MMN) in freely-moving rats with several experimental controls. PLoS ONE. doi: 10.1371/journal.pone.0110892

Hickey C, Di Lollo V, McDonald JJ (2009) Electrophysiological indices of target and distractor processing in visual search. J. Cogn. Neurosci. 21: 760-775.

Hillyard SA, Störmer VS, Feng W, Martinez A, McDonald JJ (2015) Cross-modal orienting of visual attention. Neuropsychologia doi:10.1016/j.neuropsychologia.2015.06.003

Humphreys GW, Sui J (2015) Attentional control and the self: The Self Attention Network (SAN). Cogn. Neurosci. doi:10.1080/17588928.2015.1044427

King JA, Korb FM, Egner T (2012) Priming of control: Implicit contextual cuing of top-down attentional set. J. Neurosci. 32: 8192-8200.

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

Koster EH, Crombez G, Van Damme S, Verschuere B, De Houwer J (2004) Does imminent threat capture or hold attention? Emotion 4: 312-317.

Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. 23: 571-579.

Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. 23: 571–579.

Luck SJ (2005) An Introduction to the Event-Related Potentials Technique. Cambridge, MA: MIT Press.

Mathôt S, Van der Stigchel S (2015) New light on the mind's eye: the pupillary light response as active vision. Curr. Dir. Psyc. Sci.

Matusz PJ, Eimer M (2011) Multisensory enhancement of attentional capture in visual search. Psyc. Bull. Rev. 18: 904–909.

Matusz PJ, Eimer M (2013) Top-down control of audiovisual search by bimodal search templates. Psychophysiology 50: 996–1009.

Matusz PJ, Traczyk J, Sobkow A, Strelau J (2015a) Individual differences in emotional reactivity moderate the strength of the relationship between attentional and implicit-memory biases towards threat-related stimuli. J. Cogn. Psyc. 27: 715-724.

Matusz PJ, Broadbent H, Ferrari J, Forrest B, Merkley R, Scerif G (2015b) Multi–modal distraction: Insights from children's limited attention. Cognition 136: 156–165.

Matusz PJ, Thelen A, Amrein S, Geiser E, Anken J, Murray MM (2015c) The role of auditory cortices in the retrieval of single-trial auditory–visual object memories. Eur. J. Neurosci. 41: 699-708.

McDonald JJ, Green JJ, Jannati A, Di Lollo V (2013) On the electrophysiological evidence for the capture of visual attention. J. Exp. Psyc. Hum. Perc. Perf. 39: 849–860.

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.

McDonald JJ, Green JJ, Störmer VS, Hillyard SA (2012) Cross-modal spatial cueing of attention influences visual perception. In: Frontiers in the neural bases of multisensory processes (Murray MM, Wallace, MT, eds.), pp 509–527. Boca Ratan, FL: CRC

McDonald JJ, Störmer VS, Martinez A, Feng W, Hillyard SA (2013) Salient sounds activate human visual cortex automatically. J. Neurosci. 33: 9194–201.

Meylan RV, Murray MM (2007) Auditory–visual multisensory interactions attenuate subsequent visual responses in humans. Neuroimage 35: 244-254.

Michel CM, Murray MM (2012) Towards the utilization of EEG as a brain imaging tool. Neuroimage 61: 371–385.

Moors A, De Houwer J (2006) Automaticity: A theoretical and conceptual analysis. Psyc. Bull. 132: 297–326.

Munneke J, Hoppenbrouwers SS, Theeuwes J (2015) Reward can modulate attentional capture, independent of top-down set. Att. Perc. Psychophys. DOI 10.3758/s13414-015-0958-6

Murray MM, De Santis L, Thut G, Wylie GR (2009). The costs of crossing paths and switching tasks between audition and vision. Brain Cogn. 69: 47-55.

Murray MM, Michel CM, de Peralta RG, Ortigue S, Brunet D, Andino SG, Schnider A (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. Neuroimage 21: 125-135.

Murray MM, Foxe JJ, Wylie GR (2005) The brain uses single-trial multisensory memories to discriminate without awareness. Neuroimage 27: 473-478.

Murray MM, Thelen A, Thut G, Romei V, Martuzzi R, Cappe C, Matusz PJ (2015a) The multisensory function of primary visual cortex. Neuropsychologia. In press. doi:10.1016/j.neuropsychologia.2015.08.011

Murray MM, Matusz PJ, Amedi A (2015b) Visual plasticity: Unexpected consequences of early blindness. Curr. Biol. 25. doi: 10.1016/j.cub.2015.08.054

Murray MM, Wallace, MT (2012) The Neural Bases of Multisensory Processes. Boca Ratan, FL: CRC.

Musacchia G, Schroeder CE (2009) Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. Hearing Res. 258: 72–79.

Näätänen R, Sussman ES, Salisbury D, Shafer VL (2014) Mismatch negativity (MMN) as an index of cognitive dysfunction. Brain Topogr. 27: 451-466.

Nobre K, Kastner S (2014) The Oxford Handbook of Attention. New York, NY: Oxford UP.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9: 97-113.

Perrin F, Pernier J, Bertnard O, Giard MH, Echallier JF (1987) Mapping of scalp potentials by surface spline interpolation. Electroen. Clin. Neuro. 66: 75–81.

Powers AR, Hillock AR, Wallace MT (2009) Perceptual training narrows the temporal window of multisensory binding. J. Neurosci. 29: 12265–12274.

Raij T, Ahveninen J, Lin FH, Witzel T, Jääskeläinen IP, Letham B, ... Belliveau JW (2010) Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. Eur. J. Neurosci. 31: 1772-1782.

Romei V, Murray MM, Merabet LB, Thut G (2007) Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: Implications for multisensory interactions. J. Neurosci. 27: 11465–11472.

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: 1799–805.

Rosenblum LD, Miller RM, Sanchez K (2007) Lip-read me now, hear me better later cross–modal transfer of talker–familiarity effects. Psyc. Sci. 18: 392–396.

Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. Science 274: 1926-1928.

Sawaki R, Luck SJ (2010) Capture versus suppression of attention by salient singletons: electrophysiological evidence for an automatic attend-to-me signal. Att. Perc. Psychophys. 72: 1455–1470.

Sarmiento B, Matusz PJ, Sanabria D, Murray MM (2015) Contextual factors multiplex to control multisensory processes. Hum. Brain Map. In press. doi: 10.1002/hbm.23030

Schroeder CE, Molhom S, Lakatos P, Ritter W, Foxe JJ (2004) Human–simian correspondence in the early cortical processing of multisensory cues. Cogn. Proc. 5: 140–151.

Schröger E, Bendixen A, Denham SL, Mill RW, Bőhm TM, Winkler I (2014) Predictive regularity representations in violation detection and auditory stream segregation: From conceptual to computational models. Brain Topogr. 27: 565-577.

Spence C, Driver J (2004) Crossmodal Space and Crossmodal Attention. New York, NY: Oxford UP.

Spierer L, Manuel AL, Bueti D, Murray MM (2013) Contributions of pitch and bandwidth to sound–induced enhancement of visual cortex excitability in humans. Cortex 49: 2728–2734.

Stein BE (Ed.) (2012) The New Handbook of Multisensory Processing. Cambridge, MA: MIT Press.

Stevenson RA, Wallace MT (2013) Multisensory temporal integration: Task and stimulus dependencies. Exp. Brain Res. 227: 249–261.

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. U. S. A. 106: 22456–22461.

Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends Cogn. Sci. 13: 403-409.

Tailairach J, Tournoux P (1988) Co-Planar Stereotaxic Atlas of the Human Brain. New York: Thieme Medical Publishers.

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

Ten Oever S, Schroeder CE, Poeppel D, van Atteveldt N, Zion-Golumbic E (2014) Rhythmicity and cross–modal temporal cues facilitate detection. Neuropsychologia 63: 43–50.

Ten Oever S, Romei V, van Atteveldt N, Soto-Faraco S, Murray MM, Matusz PJ (in revisions). Control mechanisms put COGs (Context-Object-Goals) into multisensory processing.

Thelen A, Cappe C, Murray MM (2012) Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. NeuroImage 62: 1478-1488.

Thelen A, Matusz PJ, Murray MM (2014) Multisensory context portends object memory. Curr. Biol. 15: R734-735.

van Atteveldt N, Murray MM, Thut G, Schroeder CE (2014) Multisensory integration: Flexible use of general operations. Neuron 81: 1240–1253.

van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. Proc. Natl. Acad. Sci. U. S. A. 102: 1181–1186.

von Kriegstein K, Giraud AL (2006) Implicit multisensory associations influence voice recognition. PLoS Biol. 4: e326.

Watson DG, Humphreys GW, Olivers CNL (2003) Visual marking: Using time in visual selection. Trends Cogn. Sci. 7: 180–186.

Wykowska A, Schubö A (2010) On the temporal relation of top–down and bottom–up mechanisms during guidance of attention. J. Cogn. Neurosci. 22: 640–654.

Wykowska A, Schubö A (2011) Irrelevant singletons in visual search do not capture attention but can produce nonspatial filtering costs. J. Cogn. Neurosci. 23: 645–660.

Wylie GR, Murray MM, Javitt DC, Foxe JJ (2009) Distinct neurophysiological mechanisms mediate mixing costs and switch costs J. Cogn. Neurosci. 21: 105-118.

Zhao J, Al-Aidroos N, Turk-Browne NB (2013) Attention is spontaneously biased toward regularities. Psyc. Sci. 24: 677–677.

Zion-Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, ... Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". Neuron 77: 980–991.

Table 1. Summary of the experimental paradigm.

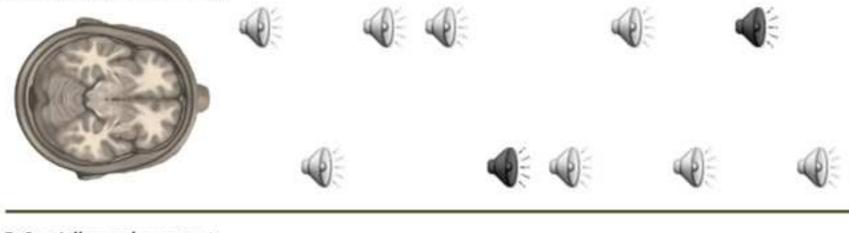| Block order (counterbalanced across subjects)	Block Type	Frequent Stimuli	Rare Stimuli
1	Spatially Irregular Context	Left-sided 250Hz (40%)	Left-sided 500Hz (10%)
		Right-sided 250Hz (40%)	Right-sided 500Hz (10%)
2	Spatially Irregular Context	Left-sided 500Hz (40%)	Left-sided 250Hz (10%)
		Right-sided 500Hz (40%)	Right-sided 250Hz (10%)
3	Spatially Regular Context	Left-sided 250Hz (40%)	Right-sided 250Hz (10%)
		Left-sided 500Hz (40%)	Right-sided 500Hz (10%)
4	Spatially Regular Context	Right-sided 250Hz (40%)	Left-sided 250Hz (10%)
		Right-sided 500Hz (40%)	Left-sided 500Hz (10%)

Figure legends

Figure 1. A schematic illustration of the experimental paradigm. **A.** The spatially irregular context involved passive listening of sounds presented with equal probability in the left and right hemispace. Darker-coloured items indicate rare trials with sounds of the same frequency (250Hz or 500Hz, counterbalanced in the experiment) that would appear in both the left and the right hemispace. **B.** The spatially regular context involved passive listening of sounds presented with 80% probability to one hemispace (here schematised for the left hemispace, though counterbalanced in the experiment). Darker-coloured items indicate rare trials with sounds of two different frequencies that would appear in one hemispace in a given block (here schematised for the right hemispace).

Figure 2. A. The upper panel displays contralateral and ipsilateral group-averaged ERPs on frequent trials, collapsed across selected occipital electrodes (the inset depicts electrodes from the contralateral region of interest). The lower panel displays the results of the 2x2 ANOVA performed on these ERPs as a function of time as well as of the corresponding planned contrasts. A reliable ACOP was observed only in response to spatially irregular sounds. **B.** The voltage topography at the latency of the peak of the ACOP (310ms) is shown for both spatially irregular and spatially regular conditions. There is a clear enhancement of occipital contralateral positive voltages in response to the sounds in the spatially irregular condition. **C.** Distributed source estimations significantly differed between the contralateral and ipsilateral responses to sounds appearing in the spatially irregular context (only clusters meeting $p < 0.05$; $k_E > 15$ nodes are shown). Stronger contralateral activity was observed in the lateral occipital cortex and precuneus (a representative axial slice is displayed). No differences were observed within auditory cortices. Nor was there any evidence of stronger ipsilateral responses. As symmetrical activity was plotted across the two hemispheres for the purposes of the source localisation, the sources observed are only plotted in the right (contralateral) hemisphere.

Figure 1

A. Spatially irregular context

B. Spatially regular context

time

A

B
Spatially irregular   Spatially regular

0.6μV

I   @310ms   C   I   @310ms   C

-0.6μV

Regularity × laterality interaction

Spatially irregular: contra vs. ipsi contrast

Spatially regular: contra vs. ipsi contrast

C
Spatially irregular
(contralateral vs. ipsilateral responses, 308-330ms)

p≤0.05
k_E>15 nodes