Contextual Factors Multiplex to Control Multisensory Processes

Beatriz R. Sarmiento,^{1,2} Pawel J. Matusz,^{3,4,5} Daniel Sanabria,^{1,2} and Micah M. Murray^{3,6,7,8*}

¹Brain, Mind and Behavior Research Center, Universidad De Granada, Spain

²Departamento De Psicología Experimental, Universidad De Granada, Spain

³The Laboratory for Investigative Neurophysiology (The LINE), Department of Radiology and Department of Clinical Neurosciences, University Hospital Centre and University of Lausanne, Lausanne, Switzerland

⁴Faculty in Wroclaw, University of Social Sciences and Humanities, Wroclaw, Poland

⁵Department of Experimental Psychology, Attention, Brain and Cognitive Development Group, University of Oxford, United Kingdom

⁶Electroencephalography Brain Mapping Core, Centre for Biomedical Imaging (CIBM), Lausanne and Geneva, Switzerland

⁷Department of Ophthalmology, University of Lausanne, Jules-Gonin Eye Hospital, Lausanne, Switzerland

⁸Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA

Abstract: This study analyzed high-density event-related potentials (ERPs) within an electrical neuroimaging framework to provide insights regarding the interaction between multisensory processes and stimulus probabilities. Specifically, we identified the spatiotemporal brain mechanisms by which the proportion of temporally congruent and task-irrelevant auditory information influences stimulus processing during a visual duration discrimination task. The spatial position (top/bottom) of the visual stimulus was indicative of how frequently the visual and auditory stimuli would be congruent in their duration (i.e., context of congruence). Stronger influences of irrelevant sound were observed when contexts associated with a high proportion of auditory-visual congruence repeated and also when contexts associated with a low proportion of congruence switched. Context of congruence and context transition resulted in weaker brain responses at 228 to 257 ms poststimulus to conditions giving rise to larger behavioral cross-modal interactions. Importantly, a control oddball task revealed that both congruent and incongruent audiovisual stimuli triggered equivalent non-linear multisensory interactions when congruence was not a relevant dimension. Collectively, these results are well explained by statistical learning, which links a particular context (here: a spatial location) with a certain level of top-down attentional control that further modulates cross-modal interactions based on whether a particular con-

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: Swiss National Science Foundation; Contract grant number: 320030-149982 (to M.M.M.); Contract grant sponsor: Ministerio de Educación y Ciencia; Contract grant numbers: AP2008-03662 (to B.R.S.) and SEJ2007-63645 (to D.S.)

*Correspondence to: Micah Murray, Department of Radiology and Department of Clinical Neurosciences, The Laboratory for Investigative Neurophysiology (The LINE), University Hospital

Centre and University of Lausanne, Lausanne, Switzerland
E-mail: micah.murray@chuv.ch

Beatriz R. Sarmiento and Pawel J. Matusz contributed equally to this work.

Received for publication 3 July 2015; Revised 2 October 2015; Accepted 5 October 2015.

DOI: 10.1002/hbm.23030

Published online 15 October 2015 in Wiley Online Library (wileyonlinelibrary.com).

text repeated or changed. The current findings shed new light on the importance of context-based control over multisensory processing, whose influences multiplex across finer and broader time scales. *Hum Brain Mapp* 37:273–288, 2016. © 2015 Wiley Periodicals, Inc.

Key words: cross-modal interaction; top-down; attention; audiovisual; context; congruence

INTRODUCTION

The interplay between top-down control processes (typically attention) and multisensory processing [e.g., Spence and Driver, 2004] is a topic of ongoing debate. Several studies have suggested that some varieties of multisensory integration occur at a preattentive stage and that it is largely impervious to attentional influences [e.g., Bertelson and de Gelder, 2004; Driver, 1996; Matusz and Eimer, 2011; reviewed in De Meo et al., 2015; Murray et al., 2015]. This notion has been challenged by other behavioral [e.g., Sanabria et al., 2007], event-related potential [ERP; e.g., Matusz and Eimer, 2013; Talsma et al., 2007], and neuroimaging [e.g., Fairhall and Macaluso, 2009] findings, supporting the claim that attention can affect multisensory outcomes [reviewed in van Atteveldt et al., 2014].

Recently, Talsma et al. [2010] proposed a framework to account for the bidirectional relationship between attention and multisensory processes. According to this model, salient multisensory stimuli are integrated before effects of attention. However, when inputs to different senses provide competing information (conflicting information may constitute an extreme example of such), top-down attention mechanisms can be activated before multisensory interactions to facilitate efficient sensory processing. For instance, Talsma and Woldorff [2005] presented evidence of an early modulation of ERPs associated with the presentation of multisensory inputs; amplitudes were enhanced when audiovisual stimuli were selectively attended. These data suggested that top-down control was needed to select appropriate (to-be-integrated) information in situations where competing sensory representations were activated.

Aside from influences based on attention (i.e., the goals of the observer), there is a growing literature emphasising the importance of control mechanisms based on context, where context can be understood as the immediate situation in which a stimulus is presented [van Atteveldt et al., 2014]. Until recently, in studies of multisensory processes these contextual influences have been studied in terms of long-term experience and learning. Studies of speech and communication signals provide clear examples of this [e.g., Froyen et al., 2009; Matusz et al., 2015a]. More recently, consideration has been given to context-based influences operating on a shorter timescale. This can include both on-line effects arising during the course of the experiment [e.g., Matusz et al., 2015b; Powers et al., 2009; Thelen et al., 2014; von Kriegstein and Giraud, 2006] as well as

effects at even finer time-scales, such as those transpiring at an inter-trial level [King et al., 2012; Murray et al., 2009; Sandhu and Dyson, 2013; Wylie et al., 2009]. For example, the fact of either repeating a task across two successive trials or switching from one task to another can dramatically influence performance, such that reaction times and error rates both increase after switching to perform a new task—i.e., the switch cost effect [reviewed in Wylie and Allport, 2000].

The importance of top-down control for stimulus processing has been broadly studied in cognitive control research using interference paradigms, where congruence between different dimensions of two unisensory stimuli is manipulated [e.g., the flanker task; Appelbaum et al., 2011; Corballis and Gratton, 2003]. Real-world environments are typically multisensory in nature and therefore congruence (or lack thereof) permeates it. For example, the congruence of low-level spatiotemporal features has been identified as a major determinant for the occurrence of multisensory integration [reviewed in Murray and Wallace, 2012; Stein, 2012]. By contrast, congruence has been largely underinvestigated with regard to the interplay of top-down control and multisensory processes.

Recently, Sarmiento et al. [2012] tested how irrelevant stimuli from one sense can interfere with the processing of stimuli in another sense by creating distinctive contexts. Participants discriminated the duration (short or long) of a visual stimulus that was accompanied by a congruent or incongruent auditory distracter. Critically, Sarmiento et al. [2012] manipulated also the likelihood of congruence between the duration of visual and auditory stimuli (hereafter context of congruence) by means of spatial locations of the auditory-visual stimuli. That is, in one spatial location (i.e., upper visual field) visual and auditory stimuli would match in their duration on the majority of trials, in contrast to the other location where on the majority of trials the stimuli would be mismatching. The difference in performance across trials containing congruent versus incongruent auditory-visual stimuli (hereafter congruence effect) was employed as an index of crossmodal interactions. The manipulation of context of congruence thus served as a proxy for understanding the importance of statistical learning for these multisensory processes [Altieri et al., 2015; Baier et al., 2006; Barakat et al., 2013; Barenholtz et al., 2014; Beierholm et al., 2009; Chandrasekaran et al., 2009].

Sarmiento et al. [2012] demonstrated that top-down attention did affect crossmodal interactions; these were larger in contexts where there was a high versus low

a. Visual Temporal Discrimination Task **b. Oddball Task** **Figure 1.**

A schematic of the experimental paradigms is displayed. **a.** The visual temporal discrimination task involved a white circle presented above or below the horizontal meridian and a white noise burst presented from two loudspeakers on each side of the computer monitor. The stimuli could either be 100 ms or 180 ms in duration and could either be congruent or incongruent in their duration across the senses. The location of the circle provided a context of congruence (i.e., the probability of

congruent or incongruent durations; see inset). **b.** The oddball task used the identical stimuli as in the visual temporal discrimination task as standard stimuli. Oddball stimuli to which participants made a button-press response were a triangle and 600 Hz tone. As above the duration of stimuli could either be 100 ms or 180 ms, and when multisensory the stimuli could either be congruent or incongruent in their duration. See Materials and Methods for full details.

probability of congruence. This was primarily due to a reduction in the congruence effect in the low-probability context. The authors argued that this situation triggered an attentional set for filtering out temporally incongruent auditory input that in turn resulted in larger crossmodal interactions. It is likely that these context-based adjustments of top-down attentional control are transpiring in an on-line fashion, as indicated by studies of purely visual incongruence [King et al., 2012]. If this holds true also for auditory-visual congruence-based interactions, such influences should also operate on a trial-to-trial basis, which can be investigated by separately quantifying crossmodal interactions (viz. congruence effects) on trials when the

context of congruence switches versus repeats. This was the principal aim of the current study.

We achieved this by using the same paradigm as in Sarmiento et al. [2012] (hereafter, visual temporal discrimination task) (Fig. 1). We expected to replicate the behavioral results of Sarmiento et al. [2012], by finding reliable influences of context on congruence-based interactions, as tested using the visual temporal discrimination task. In the present study, however, we expected that repetition of context versus a switch from one context to another on successive trials would have opposing effects on this influence. Specifically, for conditions where the context repeated we expected a larger congruence effect for the

high-probability versus low-probability context, which would not only replicate the findings of Sarmiento et al. [2012] but also extend them to show that statistical learning about contingencies between the stimulus context (i.e., a spatial location) and stimulus content can take place across single trials. By contrast, for conditions where the context switched we expected a larger congruence effect for the low-probability versus high-probability context, because the attentional set cannot be reliably established.

By including electroencephalographic measurements analysed within an electrical neuroimaging framework, we were likewise able to determine the spatiotemporal correlates of context-based effects related to statistical learning as well as those related to transitioning from one to another context from one trial to the next one. We expected to differentiate between the possible contributions of modulations in brain response strength, topography, and latency that could account for the influence of context on congruence-based auditory-visual interactions. In this way, we were likewise able to situate the effects found in the visual temporal discrimination task with respect to nonlinear multisensory integration processes as revealed by a control oddball task. In the latter, we expected to observe nonlinear response interactions between ERPs to AV stimuli and the summed responses to unisensory stimuli (i.e., A + V) [e.g., Cappe et al., 2010, 2012].

METHODS

Participants

The participants were 28 undergraduate students from the University of Granada (19 females; age range: 18–28 years old; mean age of 24 years) who received course credits or cash payment of 20 Euro in exchange of their participation. All the participants reported normal hearing and normal or corrected-to-normal vision and gave their informed consent to participate in the study conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. The data from two participants were removed due to technical issues with Net Station software. Additionally, EEG data from six participants were heavily contaminated by blink, muscle, or alpha-related artefacts, which resulted in less than 100 accepted EEG epochs for each condition of interest. Consequently, data from these subjects were removed from analyses. The data from the remaining 20 participants (12 women; age range: 18–28 years, mean age = 23 years) were included in the complete data analyses presented here.

Apparatus and Stimuli

The experiment was conducted on an Intel Core 2 Duo PC with a 17-in LCD monitor. E-Prime software was used for stimulus presentation and response collection (Psychology Software Tools; www.pstnet.com).

The visual stimuli consisted of a white circle (3.01° in diameter), a white equilateral triangle (3.01° in height) and a white central cross that served as the fixation point. The horizontal line of the cross (11.52° long) split the screen into upper and lower halves. Two loudspeakers, positioned on each side of the computer screen, were used to present the auditory stimuli, which consisted of a white noise burst and a 600 Hz pure tone (60 dB[A] measured at ear level; see Fig. 1).

Procedure

Participants sat in a comfortable chair at approximately 57 cm from the computer monitor in a silent, dark, and electrically shielded room. The instructions that explained the tasks were displayed on the computer screen. The experiment began with an unisensory discrimination task in which participants were asked to discriminate the duration (short or long) of a white circle displayed above or below $[6.52^\circ$ from the centre of the circle; cf. Experiment 4 in Sarmiento et al., 2012] the fixation point (i.e., the white horizontal line) for either 100 ms or 180 ms. The unisensory block (16 short and 16 long unisensory trials) was used to ensure that participants understood the task and were responding above 60% of accuracy overall.

After the unisensory task participants put on an elastic electrode-cap for the EEG recording (see EEG acquisition section below) during an oddball and a visual temporal discrimination task. The oddball task was conducted first in order to avoid any carryover effect from the context of congruence manipulation. However, as the visual temporal discrimination task was the main focus of the present study, we describe it first here.

Visual Temporal Discrimination Task

The oddball task was followed by the visual duration discrimination task in which participants were asked to discriminate the duration of a white circle displayed below or above the fixation point for either 100 ms or 180 ms, while ignoring the synchronous presentation of a white noise burst that could last for 100 ms or 180 ms. Onsets were always synchronous across sensory modalities. No unisensory trials were presented in the visual duration discrimination task. Congruent and incongruent trials emerged from the four possible stimuli combinations. The crucial manipulation consisted of the inclusion of two contexts of congruence defined by the two halves of the computer screen (upper/lower). For half of the participants, the upper spatial location was associated with a high proportion of congruent trials (80%) and the lower spatial location was associated with a low proportion of congruent trials (20%). The reverse was true for the remaining half of the participants. The congruent and incongruent stimuli had the same probability of appearance. This was also true for short and long trials. Consequently, the

context of congruence could repeat or switch across trials with equal probability. Notably, stimuli with identical physical properties were presented across the two contexts. This ensured that any differences observed between these contexts could be attributed to stimulus congruence probabilities being utilised by the brain to adjust top-down attentional control to facilitate visual performance by suppressing the processing of sounds in contexts with a high proportion of auditory-visual incongruence. This interpretation is in line with emerging notions regarding the interplay between top-down attention mechanisms and stimulus expectations [e.g. Aukstulewicz and Friston, 2015; Kok et al., 2012; Larsson and Smith, 2012; Summerfield and Egner, 2009].

At the beginning of each trial, participants were presented with the fixation cross displayed in white against a black background for a random duration between 500 and 1,000 ms. This fixation cross remained on throughout the whole trial. Next, the visual and auditory stimuli were presented for either 100 ms or 180 ms, and could be congruent or incongruent in duration. Feedback regarding response accuracy was provided to the participants for 500 ms. The next trial began between 1,500 and 2,000 ms after the feedback. Participants were encouraged to fixate and to avoid eye movements and blinks during the presentation of the stimuli. The sequence of events is schematized in Figure 1a.

Accuracy, rather than response speed, was emphasised; so participants had no response time pressure. Responses were collected using a serial response box. Half of the participants were told to press the leftmost key if the circle was short in duration and the rightmost key if the circle was long in duration. The reverse stimulus-response mapping was used for the remaining participants. All responses were made with the right hand. Participants completed eight multisensory practice trials and eight blocks of 120 multisensory trials each, distributed as follows: 48 congruent trials and 12 incongruent trials for the high proportion-congruent context; 12 congruent and 48 incongruent trials for the low proportion-congruent context.

Oddball Task

By means of a control, oddball task, we likewise assessed whether the context-based effects in the current experiment occur before or after any nonlinear brain response interactions between the stimuli we employed here. Importantly, we used the same stimuli as in the visual temporal discrimination task, but made participants attend to stimulus dimensions other than stimulus duration (now visual shape and sound frequency) while removing the influence of context-based control processes (Fig. 1). These manipulations enabled us to verify whether there are qualitative changes in how the congruent and incongruent stimuli from the visual temporal discrimination task are integrated when presented outside of strong

top-down influences based on task-relevance or context. A corollary question here was whether congruence would further modulate these auditory-visual interactions and, if so, at what stage. While outside of the primary focus of the current study, this important issue has thus far received limited treatment [Fort et al., 2002; Molholm et al., 2004; Yuval-Greenberg and Deouell, 2009].

Participants performed an oddball task in which either a white triangle or a 600 Hz tone had to be detected via button-press. The standard stimuli, which did not require a response, were a white circle and a white noise burst (these were identical to those in the visual temporal discrimination task described above). The visual stimuli could appear above or below the fixation cross (6.52° from the centre of the circle). Participants performed two blocks of 216 trials each, 11% of which were oddballs (i.e., 24 trials). Oddballs (triangles or 600 Hz tones) and standards (circles or noise bursts) could last for 100 or 180 ms and could be presented alone (unisensory stimuli) or synchronously with a stimulus of the other sensory modality (multisensory stimuli) with simultaneous onsets. The visual and auditory oddballs were never presented together. This manipulation gave rise to unisensory trials (the visual or auditory stimulus was presented alone), multisensory congruent trials (the visual stimulus was of the same duration as the auditory stimulus) and multisensory incongruent trials (the visual stimulus was of a different duration than the auditory stimulus). Unisensory and multisensory stimuli, short and long stimuli, as well as congruent and incongruent stimuli had the same probability of presentation (Fig. 1b). Participants had to detect the oddballs when they appeared alone as well as when they were presented accompanied by a standard stimulus from the other modality. The fixation point was presented alone for 750 or 1,250 ms, and the inter-trial interval ranged from 1,500 to 2,500 ms to avoid anticipation. Feedback was only displayed when participants made an error. The oddball task was used to assess multisensory integration between congruent and incongruent audiovisual stimuli when no response was required (i.e., standard stimuli).

At the end of the experiment, participants completed a questionnaire that evaluated whether or not they were aware of the congruence manipulation in the visual temporal discrimination task. They were asked whether they perceived any difference between the two onset locations during the visual duration discrimination task, and, if so, what that difference was.

EEG ACQUISITION

Continuous EEG was acquired at 1,000 Hz through a Geodesic Sensor Net of 129 Ag/AgCl electrodes referenced to the vertex channel [Tucker et al., 1994] and connected to an AC-coupled high-input impedance amplifier (200 M Ω , Net AmpsTM, Electrical Geodesics, Eugene, Oregon). Impedances were kept below 50 k Ω , as recommended for

these amplifiers. The signal was acquired with a 0.01 to 100 Hz elliptical band-pass filter. Gain and zero calibration were performed before the start of every recording. The head-coverage included sensors lateral to and below both eyes to monitor horizontal and vertical eye movements (electrooculogram, EOG).

ERP Analyses

The EEG was high-pass (0.1 Hz) and low-pass (30 Hz) filtered offline. Epochs of EEG were segmented from 200 ms prestimulus to 800 ms poststimulus onset. The segmented epochs were submitted to automated software processing for identification of artefacts. Epochs containing eye movements or blinks ($\pm 70 \mu\text{V}$ relative to baseline in EOG channels) were rejected. Individual channels containing other sources of transient noise ($\pm 80 \mu\text{V}$ relative to baseline in any channel) were replaced using a trial-by-trial basis with a spherical interpolation algorithm [Perrin et al., 1989]. The epoch was discarded when more than 10 channels were deemed artefact-contaminated. Additionally, the data were visually inspected to ensure that all artefacts had been detected.

Epochs of EEG were averaged for each stimulus condition and from each subject to calculate event-related potentials (ERPs). We would remind the reader that long and short duration stimuli were constituents for both congruent and incongruent stimulus conditions. By collapsing across durations, we were able to equate any contributions of offset responses to between-condition effects as well as to ensure that the comparison of responses to congruent and incongruent pairings involved equivalent stimulus energies. Data from ERPs were baseline corrected using the period of 200 ms preceding stimulus onset. ERPs were recalculated against the average reference.

For the visual temporal discrimination task, the averaging of EEG epochs was guided by the pattern of behavior. As will be detailed below, we observed a larger congruence effect for high- than low- probability contexts when context repeats and a larger congruence effect for low- than high-probability context when context switches (see Fig. 2). Consequently, we collapsed across trials with larger congruence effects as well as across trials with smaller congruence effects. This enabled us to identify brain mechanisms of the congruence effect, which serves as a proxy for audiovisual cross-modal interactions. In this way, there were four ERPs per participant following a 2×2 factorial design (congruence effect size $\times$ congruence): (1) L_{CONG} (i.e., collapsed trials when the duration was congruent and high-probability context repeated as well as when the duration was congruent and low-probability context switched), (2) L_{INCONG} (i.e., collapsed trials when the duration was incongruent and high-probability context repeated as well as when the duration was incongruent and low-probability context switched), (3) S_{CONG} (i.e., collapsed trials when the duration was congruent and low-probability context repeated as well as

Figure 2.

Accuracy on the visual temporal discrimination task displayed as the difference in percent correct performance between congruent and incongruent trials (i.e., the congruence effect). Mean congruence effects across participants (SEM indicated) are displayed as a function of context and context transition. The congruence effect was larger for the high- than for the low-probability context when the context repeated (asterisk indicates $P < 0.001$), while the reverse was true when the context switched (asterisk indicates $P < 0.001$).

when the duration was congruent and high-probability context switched), and (4) S_{INCONG} (i.e., collapsed trials when the duration was incongruent and low-probability context repeated as well as when the duration was incongruent and high-probability context switched). For each of these four conditions there were on average ($\pm$ SEM) 184(± 9), 191(± 9), 191(± 9), and 182(± 9) accepted epochs, respectively, with a minimum of 100 accepted epochs per subject and condition. Another important advantage of collapsing the data in this manner is that any differences strictly due to probability (and therefore to deviance detection) are counterbalanced across all conditions. In this way, any differences between congruent and incongruent conditions across contexts can be confidently interpreted purely in terms of top-down control.

For the oddball task, ERPs (collapsed across durations) were calculated for unisensory visual trials (V), unisensory auditory trials (A), congruent auditory-visual trials (AV_{cong}), and for incongruent auditory-visual trials (AV_{incong}). The ERPs from the A and V conditions were summed ($A + V$) and then statistically compared with the AV_{cong} and AV_{incong} conditions in a one-way nonparametric F -test. Nonlinear multisensory interactions were identified in planned contrasts (AV_{cong} vs. $A + V$ as well as AV_{incong} vs. $A + V$). Finally, planned contrasts tested for effects of congruence (AV_{cong} vs. AV_{incong}). For the AV_{cong} , AV_{incong} , and $A + V$ conditions there were on average ($\pm$ SEM) 71(± 4), 71(± 3), and 72(± 4) accepted epochs, respectively, with a minimum of 34 accepted epochs per subject and condition.

All analyses were conducted using the freeware Cartool [http://sites.google.com/site/fbmlab/cartool; Brunet et al.,

2011] and the STEN toolbox (<http://www.unil.ch/line/home/menuinst/infrastructure/software-analysis-tools.html>). The analysis strategy we used followed a multi-step analysis procedure referred to as electrical neuroimaging [Murray et al., 2008]. Electrical neuroimaging allowed us to identify effects using both local and global measures of the electric field at the scalp. This procedure distinguishes between effects following from modulations in the strength of responses of statistically indistinguishable brain generators and alterations in the configuration of these generators (viz. the topography of the electric field at the scalp), as well as latency shifts in brain processes across experimental conditions [Altieri et al., 2015; Michel and Murray, 2012; Michel et al., 2004, 2009; Murray et al., 2008; Stevenson et al., 2014]. Additionally, we used the local autoregressive average distributed linear inverse solution [LAURA; Grave de Peralta Menendez et al., 2001, 2004] to visualise and statistically contrast the likely underlying sources of effects identified in the preceding analysis steps.

Global modulations in the strength of the electric field at the scalp were quantified by calculating the global field power [GFP; Lehmann and Skrandies, 1980] for each subject and stimulus condition. This measure represents the spatial standard deviation of the electric field at the scalp at a given time point, and constitutes a reference-independent measure of the ERP amplitude [Koenig and Melie-Garcia, 2010; Murray et al., 2008]. Stronger electric fields result in larger GFP values. GFP values were baseline corrected in order to legitimately compare between conditions with different numbers of trials. GFP modulations were analyzed using a millisecond-by-millisecond nonparametric analysis of variance (ANOVA) and nonparametric planned contrasts when warranted using the Mann-Whitney test. Analyses were performed in conjunction with a 15 contiguous data-point temporal criterion for significant effects to correct for temporal auto-correlation (only those effects where P values met the statistical threshold ($P \leq 0.05$) for more than 15 contiguous time points (i.e., ~ 15 ms at 1,000 Hz sampling) were considered reliable; inspired by Guthrie and Buchwald [1991]).

In order to identify stable periods of electric field topography (hereafter template maps), the collective poststimulus group-average ERPs were subjected to hierarchical clustering. The optimal number of stable ERP clusters (i.e., the minimal number of maps that accounts for the greatest variance of the dataset) was determined using a modified Krzanowski-Lai criterion [Murray et al., 2008]. The clustering makes no assumption on the orthogonality of the derived topographic template maps [De Lucia et al., 2010a,b; Koenig et al., 2014; Pourtois et al., 2008]. Template maps identified in the group-level ERP were entered into a fitting procedure wherein each template map was compared with the moment-by-moment scalp topography on the individual subjects' ERPs from each condition and was labelled according to the one with which it best correlated spatially. This fitting procedure allows determining the

total amount of time a given template map was observed for a given condition across subjects. Statistical analysis of these values was performed with an ANOVA. The results of this analysis revealed whether a given ERP was more often described by one template map versus another, which implied different stable map configurations across conditions at the same moments in time.

Source Estimations

The intracranial generators underlying the surface electric field were estimated using a distributed linear inverse solution based on a Local Auto-Regressive Average (LAURA) model comprising biophysical laws as constraints [Grave de Peralta Menendez et al., 2001, 2004]. LAURA uses a realistic head model with a solution space of 4,024 nodes, selected from a $6 \times 6 \times 6$ mm grid equally distributed within the grey matter of the Montreal Neurological Institute's (MNI's) average brain. This algorithm selects the source configuration that better mimics the biophysical behavior of electric vector fields, that is, the estimated activity at one point depends on the activity at neighbouring points according to electromagnetic laws [Grave de Peralta and Gonzales Andino, 2002]. Before calculation of the inverse solution, the ERP data of each individual subject for each condition were down-sampled and affine-transformed to a common 111-channel montage. The time period for which intracranial sources were estimated and statistically compared between conditions was defined by the results of the abovementioned topographic analysis. Statistical analyses of source estimations were performed by first averaging the ERP data across time to generate a single data point for each participant and condition. ANOVAs were performed at each of the 4,024 source nodes in the inverse solution space (using across-subjects variance). For the temporal discrimination task, the ANOVA followed a 2×2 factorial design (congruence $\times$ congruence effect size). For the oddball task this ANOVA followed a one-way design ($AV_{\text{congr}}/AV_{\text{incongr}}/A+V$). Only nodes exceeding the spatial extent criterion of at least 17 contiguous significant nodes ($P \leq 0.0002$) were considered significant [see Thelen et al., 2012 for similar spatial criterion]. This spatial criterion was determined using the AlphaSim program (<http://afni.nimh.nih.gov>) and assuming a spatial smoothing of 6 mm full-width half maximum. AlphaSim performs 10,000 Monte Carlo permutations on the 4,024 nodes of our lead field matrix to determine the false discovery rate for clusters of different sizes. These permutations indicated that there is a 3.54% probability of a cluster of at least 17 contiguous nodes, which gives an equivalent node-level P value of $P \leq 0.0002$. The results of the analyses were rendered on the MNI average brain with loci of the largest statistical differences within a cluster indicated based on the Talairach and Tournoux [1988] atlas.

RESULTS

Visual Temporal Discrimination Task

Behavioral data

Mean response accuracy for each participant and condition were submitted to a 2 Context of congruence (high congruent, low congruent) $\times$ 2 Congruence (congruent, incongruent) $\times$ 2 Context Transition (switch, repeat) parametric ANOVA. The first trial of each block was excluded from the analysis. Neither the main effect of context transition ($F_{(1,19)} = 1.91$; $P > 0.18$; $\eta_p^2 = 0.09$) nor context of congruence ($F_{(1,19)} = 0.153$; $P = 0.7$; $\eta_p^2 < 0.01$) were reliable. There was a reliable main effect of congruence ($F_{(1,19)} = 42.55$; $P < 0.001$; $\eta_p^2 = 0.69$). None of the two-way interactions were reliable (all F 's < 1.69 ; P 's > 0.20). Crucially, there was a reliable three-way interaction ($F_{(1,19)} = 18.66$; $P < 0.001$; $\eta_p^2 = 0.49$).

Additional two-way ANOVAs for each level of the factor context transition were conducted to understand the bases for this three-way interaction. For the repeat context transition 2×2 ANOVA there was a reliable effect of congruence ($F_{(1,19)} = 36.31$; $P < 0.001$; $\eta_p^2 = 0.66$) as well as a reliable interaction between context and congruence ($F_{(1,19)} = 11.67$; $P < 0.003$; $\eta_p^2 = 0.38$). There was no main effect of context ($F_{(1,19)} = 0.36$; $P > 0.55$; $\eta_p^2 = 0.02$). The interaction on trials involving repetition of context was driven by larger in magnitude congruence effects in the high proportion-congruent context (23.8%) than in the low proportion-congruent context (16.6%); both being significant (t 's > 4.6 ; P 's < 0.001) (Fig. 2). For the switch context transition 2×2 ANOVA there was likewise a reliable effect of congruence ($F_{(1,19)} = 43.06$; $P < 0.001$; $\eta_p^2 = 0.69$) as well as a reliable interaction between context and congruence ($F_{(1,19)} = 7.06$; $P = 0.016$; $\eta_p^2 = 0.27$). There was no main effect of context ($F_{(1,19)} = 1.74$; $P > 0.20$; $\eta_p^2 = 0.08$). In contrast to the context-repeat trials, the interaction on trials involving a switch in context between trials was driven by larger in magnitude congruence effects in the low proportion-congruent context (23.3%) than in the high proportion-congruent context (17.9%); both being significant (t 's > 5.2 ; P 's < 0.001). That is, there was a larger congruence effect when switching to a low proportion-congruent context.

Participants' reports revealed that five (25%) of the participants noticed a difference between the two contexts (e.g., "I made more errors on the upper half"). However, none of them reported being aware of the specific proportion-congruent manipulation, in accordance with previous studies [Crump et al., 2006; Heinemann et al., 2009; King et al., 2012; Sarmiento et al., 2012].

Surface ERP data

In light of the behavioral results described above, the ERP data were analysed following a 2×2 Congruence effect size $\times$ Congruence factorial design (see Materials

Figure 3.

Group-averaged global field power waveforms from the visual temporal discrimination task are displayed for the L_{CONG} , L_{INCONG} , S_{CONG} , and S_{INCONG} conditions as a function of time. The yellow area plot displays the time period when a significant two-way interaction was observed (228–257 ms). The inset displays an enlarged view of the 200 to 300 ms poststimulus period. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and Methods for details). The timepoint-wise nonparametric ANOVA on the Global Field Power waveforms revealed a significant interaction over the 228 to 257 ms period (Fig. 3). Nonparametric post-hoc tests within this period showed there to be a significantly weaker GFP for L_{CONG} than L_{INCONG} over the 240 to 259 ms period. By contrast, there was a significantly stronger GFP for S_{CONG} than S_{INCONG} over the 224 to 242 ms period. Neither main effect yielded significant results.

A hierarchical topographic cluster analysis was performed on the collective group-average ERPs to identify periods of stable electric field topography both within and between experimental conditions. Eleven template maps that explained 95.9% of the variance were identified. At the group-average level there was no evidence for different template maps across conditions (see Supporting Information Fig. 1). Consequently, no single-subject fitting was performed with these data.

Source estimations

Distributed source estimations were calculated using the mean potential values over the 228 to 257 ms period and submitted to a 2 Congruence $\times$ 2 Congruence-effect size ANOVA. The main effect of congruence yielded no significant clusters. The main effect of congruence effect size yielded significant clusters centred within the left middle occipital gyrus, the right parahippocampal gyrus, the right precuneus, and the right inferior frontal gyrus. These regions are consistent with the literature reporting their contribution to cross-modal and multisensory processes

[reviewed in Murray and Wallace, 2012], including those focusing on stimulus (a)synchrony [e.g., Macaluso et al., 2004]. A significant interaction between congruence and congruence-effect size was observed in a widely distributed network of clusters within the left superior temporal cortices, the left parietal cortices, and bilateral precuneus (Fig. 4a). In particular, the parietal cortices likely included pre-motor structures previously implicated in temporal processes [e.g., Grahm and Rowe, 2009; Marchant et al., 2012]. Given this interaction, contrasts focused on the congruence effect for the large-size congruence effect (L_{CONG} vs. L_{INCONG}) and small-size congruence effect (S_{CONG} vs. S_{INCONG}) conditions, separately. For the large-size congruence effect conditions, clusters exhibiting significant differences were centred within the left medial frontal gyrus, the left superior temporal cortex, and the precuneus bilaterally (Fig. 4b). All clusters were more strongly active in response to incongruent than congruent trials. For the small-size congruence effect conditions, clusters exhibiting stronger responses to congruent than incongruent pairings were centred within the left parietal cortices (Fig. 4c).

Oddball Task

Behavioral data

Overall accuracy across participants on the oddball task was 99.77%, with a mean reaction time of 340 ms. This high level of performance provided an assurance that participants were appropriately engaged in the oddball task.

Surface ERP data

Analyses of data from the Oddball task focused on determining differences in the neural activation between the ERPs in response to the congruent and incongruent multisensory conditions and the summed ERPs in response to the unisensory auditory and visual conditions (denoted by AV_{cong} , AV_{incong} and $A + V$, respectively). Only trials for standard stimuli were included in the analyses, because they constituted the pool of stimuli used for the visual temporal discrimination task, the analyses of which are described above.

The GFP waveforms were compared statistically using a one-way ANOVA as a function of time (Fig. 5). There was a significant main effect over two time intervals (130–145 ms and 249–800 ms). Follow-up planned nonparametric contrasts revealed no significant differences between AV_{cong} and AV_{incong} responses. Likewise, for both time windows, there were super-additive multisensory responses that were observed irrespective of stimulus congruence.

Next, a hierarchical spatiotemporal cluster analysis was employed again to test for topographic differences between experimental conditions within the data from the oddball task. This procedure identified a set of 11 template maps explaining 96.3% of the variance of the concatenated group-averaged ERP data set. There was no evidence for

different template maps across conditions (see Supporting Information Fig. 2). Consequently, no single-subject fitting was performed with these data.

Source estimations

Estimated intracranial sources over the 130 to 145 ms poststimulus time period revealed a main effect of stimulus condition centred within a distributed network of areas including the calcarine sulcus bilaterally, the left parieto-occipital cortex and left parietal cortices, the orbitofrontal cortex bilaterally and the right lateral occipital cortices (see Fig. 6a). These areas have been repeatedly observed in studies of audiovisual integration [e.g., Calvert et al., 2000; Cappe et al., 2010, 2012; Dhamala et al., 2007], with the orbitofrontal cortices implicated in multisensory conflict [Diaconescu et al., 2011]. Planned comparisons between the AV_{cong} and $A + V$ conditions revealed super-additive responses within the calcarine sulcus bilaterally, the left parieto-occipital cortex and left parietal cortices, and sub-additive responses within the orbitofrontal cortex bilaterally (Fig. 6b). A similar pattern of effects was observed between AV_{incong} and $A + V$, with the addition of super-additive responses within the right lateral occipital cortices (Fig. 6c). Finally, contrasts between AV_{cong} and AV_{incong} revealed stronger responses to the AV_{cong} condition within the calcarine sulcus bilaterally and stronger responses to the AV_{incong} condition within the right lateral occipital cortex, the right parietal cortex and the bilateral orbitofrontal cortex (Fig. 6d). However, we are hesitant to over-interpret these results, given that corresponding effects were not reliably at the level of scalp ERPs.

Source estimations over the 195 to 275 ms poststimulus time period revealed a main effect of stimulus condition throughout an extensive brain network including bilateral occipital, parietal, temporal, and frontal sources (Supporting Information Fig. 3). Planned comparisons between either AV_{cong} and $A + V$ or AV_{incong} and $A + V$ revealed subadditive responses within this network. The contrast between AV_{cong} and AV_{incong} revealed stronger responses to AV_{cong} within the left temporal pole, the anterior cingulate, and dorsal prefrontal cortex. Stronger responses to the AV_{incong} condition were observed in right parietal cortices. This locus is similar to that observed for the large-size congruence effect contrast (see Fig. 4b).

DISCUSSION

It is well established that some multisensory processes can occur in a bottom-up fashion, i.e., independently of top-down control based on one's goals. Much less is known about how these processes are controlled by influences based on the context in which the multisensory stimuli are presented as well as whether (and if so how) context-based mechanisms operating across different time-scales interact with each other.

a. 2 Context of congruence x2 Congruence-effect size interaction (228-257 ms)

b. Large-size congruence-effect contrast

c. Small-size congruence-effect contrast

Figure 4.

The results of statistical analyses of distributed source estimations from the visual temporal discrimination task are displayed. **a.** The 2×2 interaction yielded significant clusters ($P < 0.05$; $k_E > 17$ contiguous nodes) centred within the left middle occipital gyrus, the right parahippocampal gyrus, the right precuneus, and the right inferior frontal gyrus. **b.** Follow-up planned contrasts for the large-size congruence-effect revealed significant

clusters centred within the left medial frontal gyrus, the left superior temporal cortex, and the precuneus bilaterally. All clusters were more strongly active in response to incongruent than congruent trials. **c.** Follow-up planned contrasts for the small-size congruence-effect revealed significant clusters centred within the left parietal cortices with stronger responses to congruence than incongruent pairings.

Figure 5.

Group-averaged global field power waveforms from the oddball task are displayed for the AV_{congr}, AV_{incongr}, and A + V conditions as a function of time. The red area plot displays the time periods when a significant one-way interaction was observed (130–145 and 249–800 ms; truncated in the figure at 500 ms). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

In a series of behavioral experiments, Sarmiento et al. [2012] demonstrated that the ability of an irrelevant auditory stimulus to interfere with a duration judgement on a visual stimulus can be influenced by the particular context in which the stimuli appear. Specifically, spatial locations where there was a large proportion of congruent auditory-visual stimuli were generally giving rise to stronger influences of the sound on visual processing than in spatial locations with higher proportion of incongruent auditory-visual presentations. The authors interpreted these findings in terms of specific contexts activating top-down attentional control mechanisms to a different extent. Contexts with a high proportion of incongruent pairings trigger enhanced top-down inhibition that in turn attenuated auditory-visual interactions, indexed by behavioral congruence effects. This top-down control enhancement likely occurred implicitly, as indicated by the majority of participants being unaware of the experimental manipulation.

The present study employed the same paradigm to understand whether context-based control operates at a finer ‘online’ time scale. To do this, we analysed trials when the context of congruence repeated versus switched on two successive trials, and electrophysiological recordings were used to understand the brain mechanisms underlying the influence of these control mechanisms. Furthermore, we added a control, oddball task, where the same stimuli as those used on the visual temporal discrimination task were both made irrelevant, so that the timing of the nonlinear interactions between these stimuli could be assessed outside of influences of top-down control and then compared with the timing of the occurrence of context-based influences described above.

The visual temporal discrimination task revealed that congruence effects differed in magnitude depending on the context of congruence as well as on whether a particular context repeated or switched. Specifically, larger-size congruence effects were observed for high-proportion congruence contexts on trials where the context repeated and for low-proportion congruence contexts on trials where the context switched (Fig. 2). Our analyses of the brain responses focused on the mechanisms giving rise to large versus small congruence effects at the level of behavior. Audiovisual interactions were modulated by context transition at ~230 ms poststimulus, with these modulations being driven by changes in brain response strength within a statistically indistinguishable configuration of brain areas (Fig. 3). In the case of large-size congruence effects, brain responses were attenuated on trials with congruent than incongruent auditory-visual stimuli, and originating in the left medial frontal gyrus, the left superior temporal cortex, and the precuneus bilaterally (Fig. 4). Small-size congruence effects were found to be associated with stronger responses on congruent than incongruent trials in the left parietal cortices.

Auditory-visual interactions in the visual temporal discrimination task differed in their strength depending on which context they appeared in as well as whether this context repeated or switched across two successive trials. Larger (and comparable in size) congruence effects were observed in contexts with high proportion of auditory-visual congruence on trials where this context repeated and in contexts with high proportion of cross-modal *incongruence* when this context *switched*. These results can be most readily interpreted in terms of the stimulus context (or rather contexts operating both at coarser and finer time scales) modulating the cross-modal interactions by means of a form of statistical learning.

Statistical learning is typically understood as a process by which a cognitive system learns about the underlying structure of the sensory inputs by extracting the distributional properties from these inputs across time and/or space [Saffran et al., 1996]. While the domain-general versus -specific nature of this process is a topic of current debate [Frost et al., 2015], it has been demonstrated to support a variety of cognitive functions, from auditory stimulus parsing and visual search to conditioning, to name just a few [e.g., Baker et al., 2004; Courville et al., 2006; Goujon and Fagot, 2013; Saffran et al., 1996]. The importance of statistical learning has likewise been demonstrated for multisensory processing [Altieri et al., 2015; Baier et al., 2006; Barakat et al., 2013; Barenholtz et al., 2014; Beierholm et al., 2009; Chandrasekaran et al., 2009]. For example, one can learn to associate arbitrary but spatially and temporally congruent auditory and visual stimuli (pure tones and Gabor patches, respectively), with reduced brain response strength corroborating increased efficiency of their recognition after a few days of training [Altieri et al., 2015]. In a similar vein, learning of pairs consisting of

Figure 6.

The results of statistical analyses of distributed source estimations over the 130 to 145 ms time period from the oddball task are displayed. **a.** The one-way ANOVA yielded significant clusters ($P < 0.05$; $k_E > 17$ contiguous nodes) centred within a distributed network of areas including the calcarine sulcus bilaterally, the left parieto-occipital cortex and left parietal cortices, the orbitofrontal cortex bilaterally and the right lateral occipital cortices. **b.** Follow-up planned contrasts between AV_{cong} and $A + V$ revealed super-additive responses within the calcarine sulcus bilaterally, the left parieto-occipital cortex and left parietal

cortices, and sub-additive responses within the orbitofrontal cortex bilaterally. **c.** A similar pattern of effects was observed between AV_{incong} and $A + V$, with the addition of super-additive responses within the right lateral occipital cortices. **d.** The contrast between AV_{cong} and AV_{incong} revealed stronger responses to the AV_{cong} condition within the calcarine sulcus bilaterally and stronger responses to the AV_{incong} condition within the right lateral occipital cortex, the right parietal cortex and the orbitofrontal cortex (bilaterally). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

human voices and gender-congruent faces is facilitated compared to pairs involving human voices and images of gender-incongruent faces or of plants/rocks [Barenholtz et al., 2014].

How do the current results fit in with our understanding of the role of statistical learning and context in control-

ling multisensory processing? Top-down attentional control would be a prime candidate for a mechanism that could link a particular location, a particular audiovisual stimulus (congruent or incongruent) and, crucially, a particular context of congruence [Treisman, 1996, 2006]. Indeed, research on implicit learning has repeatedly

shown that selective attention is critical in order to establish a link between predictive dimensions of a particular stimulus (probability of congruence and spatial location in this case), given that such associations only emerge when these dimensions are relevant to the observer [Crump et al., 2008; Jiménez and Méndez, 1999; see also Auksztulewicz and Friston, 2015; Kok et al., 2012; Larsson and Smith, 2012 for evidence on interactions between top-down attention and expectations]. In the present study, the location was crucial since the participant needed to localize the target (displayed in one of two possible locations) in order to perform the task. Once the association between proportion of congruence and location was established, top-down attentional control of differing strength was applied to perform the task within the particular context [see e.g. Vossel et al., 2014 for a recent study modeling saccadic response—times data from a spatial-attention task]. As a result, the very same stimulus (e.g., congruent) was processed differently as a function of the location where it was presented (that determined the context of congruence) and whether this context had repeated or switched. An additional contributing factor may therefore be whether or not the stimulus itself was repeated or switched (in addition to whether the context itself repeated or switched). Unfortunately, in the present study, it was infeasible to include this additional factor, principally because it would have severely reduced the signal-to-noise ratio of the ERP data. Notwithstanding, stimulus repetition effects in the current paradigm would be a particularly interesting avenue to pursue in future research, given recent evidence for the interplay between stimulus repetition and factors such as expectancy, attention, and memory [e.g., Doehrmann et al., 2010; Henson et al., 2000; Recasens et al. 2014; Turk-Browne et al., 2007; reviewed in Segaert et al., 2013].

In a functional neuroimaging study employing a visual face-gender classification version of the flanker task involving a location-based contextual manipulation of conflict frequency, King et al. [2012] demonstrated that reduced interference in contexts with high proportion of conflict is associated with stronger engagement of top-down control areas (predominantly the medial superior parietal lobule), with their comparatively weak engagement in contexts with a low proportion of conflict. The current results are in line with differing engagement of top-down attentional control across the two contexts of congruence. If a particular location becomes associated with a high proportion of congruence between the visual and auditory stimuli, there is no need for heightened top-down control as the presence of the irrelevant sounds would typically improve the visual duration discrimination. The interactions arising between the auditory and visual stimuli are likely to be larger in the case of a repeated context that involves weak top-down inhibition of the irrelevant sounds. Along the same lines, a repeated context involving stronger top-down inhibition, as in the

case of the context involving a low proportion of cross-modal congruence, is bound to lead to attenuated auditory-visual interactions when compared with a switch within such a context.

Several important aspects of the current results support the interpretation in terms of online and implicit adjustment of top-down attentional control based on a particular context of congruence as well as whether it repeated or switched. First, we hasten to remind the reader that the stimuli appearing in the two contexts were identical in their physical properties (i.e., same number of auditory-visual congruent and incongruent trials). This excludes the possibility that differences in brain responses were due to differences in stimulation or stimulus-response associations for each context. Second, by collapsing the ERP across conditions involving high-probability and low-probability contexts we also have minimised any confounding influences on the brain activity stemming strictly from detection of improbable events [e.g., P300, whose latency would coincide with the latency of our effects; reviewed in Polich, 2007]. Thus, the observed ERP differences can reasonably be interpreted in terms of the brain utilising the probabilities with which stimuli were congruent in particular contexts to adjust top-down attentional control. Lastly, the lack of explicit awareness in the majority of participants regarding the congruence manipulation in the visual temporal discrimination task accords with previous findings using similar paradigms [e.g., Crump et al., 2008; King et al., 2012; Heinemann et al., 2009; Sarmiento et al., 2012]. This result suggests that the top-down attention adjustments were likely occurring outside of the voluntary control of participants. This notion is further reinforced by the differences observed between repeated and switched contexts, indicating the online (i.e., trial-by-trial) nature of these implicit top-down attention adjustments.

The current study demonstrates how important for cross-modal interactions is the context in which multisensory stimuli appear, with its effects multiplexing across finer and broader time scales to influence these interactions. A predominance of congruence between signals across different senses will determine the level of top-down inhibition the brain associates with a particular context. This will have a cascading effect, with stronger cross-modal interactions transpiring across contexts associated with both high and low proportion of auditory-visual congruence, on trial-to-trial variations in the magnitude of cross-modal interactions [Murray et al., 2009; Sandhu and Dyson, 2013]. Importantly, the current results point to the fact that incongruity does not necessarily trigger heightened top-down control. More broadly, the current study sheds new light on the ongoing debate regarding the influence of top-down control on multisensory processes [De Meo et al., 2015; Murray et al., 2015; van Atteveldt et al., 2014; Talsma et al., 2010; ten Oever et al., unpublished data].

The employment of a control, oddball task enabled us to compare the timing of nonlinear multisensory interactions

with that of cross-modal interactions elicited by the same stimuli when under top-down control. Nonlinear multisensory interactions first occurred in the present dataset at 130 to 145 ms poststimulus, independent of the stimulus congruence. By contrast, effects in the visual temporal discrimination task occurred first at ~230 ms poststimulus (i.e. some 100 ms later). While the bases for this latency difference remain to be fully identified, these results corroborate an emerging consensus that timing of multisensory interactions at the brain level might be one way of distinguishing processes less and more dependent on top-down control [reviewed in De Meo et al., 2015]. Some multisensory processes, such as those based on simultaneity, seem to occur independently of top-down attention and context influences, being observed across paradigms and task demands, as well as across different species. These bottom-up, early multisensory interactions (eMSI) are typically observed within 100 ms poststimulus onset [De Meo et al., 2015]. Some of the strongest support for their independence of top-down attention or context is provided by studies reporting eMSI in anaesthetised preparations [reviewed in reviews in Rowland and Stein, 2014; Sarko et al., 2012]. In contrast, multisensory processes based on congruence [e.g., perceptual as in the current study; see also Fort et al., 2002; semantic in Molholm et al., 2004; Yuval-Greenberg and Deouell, 2009] typically transpire at later poststimulus stages. In the case of the present study, the timing differences across the two tasks may also be linked to the fact that the visual temporal discrimination can at the earliest be completed only after 100 ms poststimulus (i.e. at the duration of the shorter stimulus). Future work will need to either modify the task to allow for its potential completion upon stimulus onset or otherwise manipulate parametrically the said timing to more fully dissociate multisensory effects differently dependent on top-down attention. Collectively, the documentation of multisensory effects across multiple time scales corroborate a novel framework differentiating top-down control processes based on attention (i.e., goals), context and content (ten Oever et al., in revision). The current results support this proposal by showing that both congruent and incongruent stimuli triggered nonlinear multisensory response interactions at ~140 ms poststimulus; in some instances stimulus congruence might not play a role in modulating multisensory processing unless it is relevant to the current task [see also Molholm et al., 2004, for similar multisensory findings; see Matusz et al., in revision, for cross-modal results on predictability enabling the suppression of task-irrelevant sounds].

CONCLUSIONS

In a visual temporal discrimination task, stronger influences of irrelevant sound were observed in repeated contexts associated with a high proportion of auditory-visual congruence and following switch in context associated

with a low proportion of congruence. Multisensory processes are therefore subject to control both by context as well as by statistical learning, where the latter operates on at least two time scales, but only in situations where congruence across the senses is relevant for task demands. These findings provide important insights into the interactions between expectations and top-down attention occurring in environments closely resembling naturalistic ones, i.e. where stimuli differ in their relevance to the current goals of the observer as well as the sensory modality in which they appear.

ACKNOWLEDGMENT

The authors thank Diana Cutanda for her help with data collection and Antonia Thelen for her assistance with data analyses.

REFERENCES

- Altieri N, Stevenson RA, Wallace MT, Wenger MJ (2015): Learning to associate auditory and visual stimuli: Behavioral and neural mechanisms. *Brain Topogr* 28:479–493.
- Appelbaum LG, Smith DV, Boehler CN, Chen WD, Woldorff MG (2011): Rapid modulation of sensory processing induced by stimulus conflict. *J Cogn Neurosci* 23:2620–2628.
- Auksztulewicz R, Friston K (2015): Attentional enhancement of auditory mismatch responses: A DCM/MEG Study. *Cereb Cortex* 1–11. (Epub ahead of print). doi:10.1093/cercor/bhu323.
- Baier B, Kleinschmidt A, Müller NG (2006): Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. *J Neurosci* 26:12260–12265.
- Baker CI, Olson CR, Behrmann M (2004): Role of attention and perceptual grouping in visual statistical learning. *Psychol Sci* 15:460–466.
- Barakat BK, Seitz AR, Shams L (2013): The effect of statistical learning on internal stimulus representations: Predictable 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. *Psychol 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.
- Bertelson P, de Gelder B (2004): The psychology of multimodal perception. In: Spence C, Driver J, editors. *Crossmodal Space and Crossmodal Attention*. Oxford: Oxford University Press. pp 141–177.
- Brunet D, Murray MM, Michel CM (2011): Spatiotemporal analysis of multichannel EEG: CARTOOL. *Comput Intell Neurosci* 2011:813–870.
- Calvert GA, Campbell R, Brammer MJ (2000): Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649–657.
- Cappe C, Thut G, Romei V, Murray MM (2010): Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J Neurosci* 30:12572–12580.

- Cappe C, Thelen A, Romei V, Thut G, Murray MM (2012): Looming signals reveal synergistic principle of multisensory integration. *J Neurosci* 32:1171–1182.
- Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA (2009): The natural statistics of audiovisual speech. *PLoS Comput Biol* 5:e1000436.
- Corballis PM, Gratton G (2003): Independent control of processing strategies for different locations in the visual field. *Biol Psychol* 64:191–209.
- Courville AC, Daw ND, Touretzky DS (2006): Bayesian theories of conditioning in a changing world. *Trends Cogn Sci* 10:294–300.
- Crump MJ, Gong Z, Milliken B (2006): The context-specific proportion congruent stroop effect: Location as a contextual cue. *Psychon Bull Rev* 13:316–321.
- Crump MJC, Vaquero JMM, Milliken B (2008): Context-specific learning and control: The role of awareness, task relevance, and relative salience. *Conscious Cogn* 17:22–36.
- De Lucia M, Clarke S, Murray MM (2010a): A temporal hierarchy for conspecific vocalization discrimination in humans. *J Neurosci* 30: 11210–11221.
- De Lucia M, Michel CM, Murray MM (2010b): Comparing ICA-based and single-trial topographic ERP analyses. *Brain Topogr* 23: 119–127.
- 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.
- Dhamala M, Assisi CG, Jirsa VK, Steinberg FL, Kelso JAS (2007): Multisensory integration for timing engages different brain networks. *Neuroimage* 34:764–773.
- Diaconescu AO, Alain C, McIntosh AR (2011): The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *J Neurophys* 106:2896–2909.
- Doehrmann O, Weigelt S, Altmann CF, Kaiser J, Naumer MJ (2010): Audiovisual functional magnetic resonance imaging adaptation reveals multisensory integration effects in object-related sensory cortices. *J Neurosci* 30:3370–3379.
- Driver J (1996): Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* 381: 66–68.
- Fairhall SL, Macaluso E (2009): Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *Eur J Neurosci* 29:1247–1257.
- Fort A, Delpuech C, Pernier J, Giard MH (2002): Dynamics of cortico-subcortical cross-modal operations involved in audiovisual object detection in humans. *Cereb Cortex* 12: 1031–1039.
- Frost R, Armstrong BC, Siegelman N, Christiansen MH (2015): Domain generality versus modality specificity: The paradox of statistical learning. *Trends Cogn Sci* 19:117–125.
- Froyen D, van Atteveldt N, Bonte M, Blomert L (2009): Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neurosci Lett* 430:23–28.
- Goujon A, Fagot J (2013): Learning of spatial statistics in nonhuman primates: Contextual cueing in baboons (*Papio papio*). *Behav Brain Res* 247:101–109.
- Grahn JA, Rowe JB (2009): Feeling the beat: Premotor and striatal interactions in musicians and non-musicians during beat perception. *J Neurosci* 29:7540–7548.
- Grave de Peralta Menendez R, Gonzalez Andino S, 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, R, Gonzalez Andino S (2002): Comparison of algorithms for the localization of focal sources: Evaluation with simulated data and analysis of experimental data. *Int J Bioelectromagn* 4:1.
- Grave de Peralta Menendez R, Murray MM, Michel CM, Martuzzi R, Gonzalez Andino SL (2004): Electrical neuroimaging based on biophysical constraints. *Neuroimage* 21: 527–539.
- Guthrie D, Buchwald JS (1991): Significance testing of difference potentials. *Psychophysiology* 28: 240–244.
- Heinemann A, Kunde W, Kiesel A (2009): Context-specific prime-congruency effects: On the role of conscious stimulus representations for cognitive control. *Conscious Cogn* 18:966–976.
- Henson R, Shallice T, Dolan R (2000): Neuroimaging evidence for dissociable form of repetition priming. *Science* 287:1269–1272.
- Jiménez L, Méndez C (1999): Which attention is needed for implicit sequence learning? *J Exp Psychol Learn Mem* 25:236–259.
- King JA, Korb FM, Egner T (2012): Priming of control: Implicit contextual cuing of top-down attentional set. *J Neurosci* 32: 8192–8200.
- Koenig T, Melie-Garcia L (2010): A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topogr* 23: 233–242.
- Koenig T, Stein M, Grieder M, Kottlow M (2014): A tutorial on data driven methods for statistically assessing ERP topographies. *Brain Topogr* 27:72–83.
- Kok P, Rahnev D, Jehee JFM, Lau HC, de Lange FP (2012): Attention reverses the effect of prediction in silencing sensory signals. *Cereb Cortex* 22:2197–2206.
- Larsson J, Smith AT (2012): fMRI repetition suppression: Neuronal adaptation or stimulus expectation? *Cereb Cortex* 22:567–576.
- Lehmann D, Skrandies W (1980): Reference-free identification of components of checkerboard evoked multichannel potential fields. *Electroencephalogr Clin Neurol* 48:609–621.
- Macaluso E, George N, Dolan R, Spence C, Driver J (2004): Spatial and temporal factors during processing of audiovisual speech: A PET Study. *Neuroimage* 21:725–732.
- Marchant JL, Ruff CC, Driver J (2012): Audiovisual synchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target-detection performance for both modalities. *Hum Brain Mapp* 33:1212–1224.
- Matusz PJ, Broadbent H, Ferrari J, Forrest B, Merkley R, Scerif G (2015a): 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 (2015b): The role of auditory cortices in the retrieval of single-trial auditory-visual object memories. *Eur J Neurosci* 41: 699–708.
- Matusz PJ, Eimer M (2013): Top-down control of audiovisual search by bimodal search templates. *Psychophysiology* 50:996–1009.
- Matusz PJ, Eimer M (2011): Multisensory enhancement of visual attentional capture in visual search. *Psychon Bull Rev* 18:904–909.
- Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, Grave de Peralta R (2004): EEG source imaging. *Clin Neurophysiol* 115:2195–2222.
- Michel CM, Koenig T, Brandeis D, Gianotti LR, Wackermann J (2009): Electrical Neuroimaging. Cambridge, UK: Cambridge UP.
- Michel CM, Murray MM (2012): Towards the utilization of EEG as a brain imaging tool. *Neuroimage* 61:371–385.
- Molholm S, Ritter W, Javitt DC, Foxe JJ (2004): Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cereb. Cortex* 14:452–465.

- Murray MM, Brunet D, Michel CM (2008): Topographic ERP analyses: A step-by-step tutorial review. *Brain Topogr* 20:249–264.
- 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, Wallace MT (2012): *The Neural Basis of Multisensory Processes*. Boca Raton, FL: CRC Press.
- Murray MM, Thelen A, Thut G, Romei V, Martuzzi R, Matusz PJ (2015): The multisensory function of the human primary visual cortex. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2015.08.011.
- Perrin F, Pernier J, Bertrand O, Echallier JF (1989): Spherical splines for scalp potential and current density mapping. *Electroen clin Neurol* 72, 184–187.
- Powers AR, Hillock AR, Wallace MT (2009): Perceptual training narrows the temporal window of multisensory binding. *J Neurosci* 29:12265–12274.
- Polich J (2007): Updating P300: An integrative theory of P3a and P3b. *Clin Neurophys* 118:2128–2148.
- Pourtois G, Delplanque S, Michel C, Vuilleumier P (2008): Beyond conventional event-related brain potential (ERP): Exploring the time-course of visual emotion processing using topographic and principal component analyses. *Brain Topogr* 20: 265–277.
- Recasens M, Leung S, Grimm S, Nowak R, Escera C (2014): Repetition suppression and repetition enhancement underlie auditory memory-trace formation in the human brain: An MEG study. *Neuroimage* 108:75–86.
- Rowland BA, Stein BE (2014): A model of the temporal dynamics of multisensory enhancement. *Neurosci Biobehav Rev* 41:78–84.
- Saffran JR, Aslin RN, Newport EL (1996): Statistical learning by 8-month-old infants. *Science* 274:1926–1928.
- Sanabria D, Soto-Faraco S, Spence C (2007): Spatial attention and audiovisual interactions in apparent motion. *J Exp Psychol Hum Percept Perform* 33: 927–937.
- Sanabria D, Spence C, Soto-Faraco S (2007): Perceptual and decisional contributions to audiovisual interactions in the perception of apparent motion: A signal detection study. *Cognition* 102:299–310.
- Sandhu R, Dyson BJ (2013): Modality and task switching interactions using bi-modal and bivalent stimuli. *Brain Cogn* 82:90–99.
- Sarko DK, Nidiffer AR, Powers AR III, Ghose D, Wallace MT (2012): In: Murray MM, Wallace MT, editors. *The Neural Basis of Multisensory Processes*. Boca Raton, FL: CRC Press. pp 191–215.
- Sarmiento BR, Shore D, Milliken B, Sanabria D (2012): Audiovisual interactions depend on context of congruency. *Attention Percept Psychophys* 74:563–574.
- Segaert K, Weber K, de Lange FP, Petersson KM, Hagoort P (2013): The suppression of repetition enhancement: A review of fMRI studies. *Neuropsychologia* 51:59–66.
- Spence C, Driver J, editors (2004): *Crossmodal Space and Crossmodal Attention*. Oxford, UK: Oxford UP.
- Stein BE (2012): *The New Handbook of Multisensory Processing*. MIT Press: Cambridge, MA.
- Stevenson RA, Ghose D, Fister JK, Sarko DK, Altieri NA, Nidiffer AR, Kurela LR, Siemann JK, James TW, Wallace MT (2014): Identifying and quantifying multisensory integration: A tutorial review. *Brain Topogr* 27:707–730.
- Summerfield C, Egner T (2009): Expectation (and attention) in visual cognition. *Trends Cogn Sci* 13:403–409.
- Talairach J, Tournoux P (1988): *Co-planar stereotaxic atlas of the human brain: 3- dimensional proportional system: An approach to cerebral imaging*. G. Thieme Medical Publishers: New York.
- Talsma D, Woldorff MG (2005): Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *J Cogn Neurosci* 17: 1098–1114.
- Talsma D, Doty TJ, Woldorff MG (2007): Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cereb Cortex* 17:679–690.
- Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010): The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci* 14:400–410.
- 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–R735.
- Treisman A (1996): The binding problem. *Curr Opin Neurobiol* 6: 171–178. doi: 10.1016/S0959-4388(96)80070-5
- Treisman A (2006): How the deployment of attention determines what we see. *Vis Cogn* 14: 411–443.
- Tucker DM, Liotti M, Potts GF, Russell GS, Posner MI (1994): Mapping the working brain: Spatiotemporal analysis of brain electrical fields. *Hum Brain Mapp* 1:134–152.
- Turk-Browne NB, Yi DJ, Leber AB, Chun MM (2007): Visual quality determines the direction of neural repetition effects. *Cereb Cortex* 17:425–433.
- van Atteveldt N, Murray MM, Thut G, Schroeder CE (2014): Multisensory integration: Flexible use of general operations. *Neuron* 81:1240–1253.
- Von Kriegstein K, Giraud AL (2006): Implicit multisensory associations influence voice recognition. *PLoS Biol* 4:e326.
- Vossel S, Mathys C, Daunizeau J, Bauer M, Driver J, Friston KJ, Stephan KE (2014): Spatial attention, precision, and Bayesian inference: A study of saccadic response speed. *Cereb Cortex* 24:1436–1450.
- Wylie G, Allport A (2000): Task switching and the measurement of “switch costs”. *Psychol Res* 63:212–233.
- Wylie GR, Murray MM, Javitt DC, Foxe JJ (2009): Distinct neurophysiological mechanisms mediate mixing costs and switch costs. *J Cogn Neurosci* 21:105–118.
- Yuval-Greenberg S, Deouell LY (2009): The dog’s meow: Asymmetrical interaction in cross-modal object recognition. *Exp Brain Res* 193:603–614.