

Serveur Académique Lausannois SERVAL serval.unil.ch

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:

Title: Contextual factors multiplex to control multisensory processes.

Authors: Sarmiento BR, Matusz PJ, Sanabria D, Murray MM

Journal: Human brain mapping

Year: 2016 Jan

Issue: 37

Volume: 1

Pages: 273-88

DOI: 10.1002/hbm.23030

In the absence of a copyright statement, users should assume that standard copyright protection applies, unless the article contains an explicit statement to the contrary. In case of doubt, contact the journal publisher to verify the copyright status of an article.



Contextual factors multiplex to control multisensory processes

Journal:	<i>Human Brain Mapping</i>
Manuscript ID	HBM-15-0803.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	02-Oct-2015
Complete List of Authors:	Sarmiento, Beatriz; Universidad de Granada, Departamento de Psicología Experimental Matusz, Pawel; Centre Hospitalier Universitaire Vaudois and University of Lausanne, Department of Clinical Neurosciences Sanabria, Daniel; Universidad de Granada, Dept. of Psicología Experimental. Mind, Brain and Behavior research center Murray, Micah; Centre Hospitalier Universitaire Vaudois and University of Lausanne, EEG Brain Mapping Core, Center for Biomedical Imaging, Neuropsychology and Neurorehabilitation, Radiology
Keywords:	cross-modal interaction, multisensory, audiovisual, ERP, context, congruence

SCHOLARONE™
Manuscripts

RUNNING HEAD: *Multiplexed control over multisensory processes*

Contextual factors multiplex to control multisensory processes

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

¹Departamento de Psicología Experimental, Universidad de Granada, Spain

²Brain, Mind and Behavior Research Center, 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

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

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

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

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

*equal contributions

Correspondence to:

Prof. Micah Murray (micah.murray@chuv.ch)

University Hospital Center and University of Lausanne

Abstract

This study analysed 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 spatio-temporal 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 228-257 ms post-stimulus to conditions giving rise to larger behavioural 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 context 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.

Keywords: 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 pre-attentive stage and that it is largely impervious to attentional influences (e.g., Driver, 1996; Bertelson and de Gelder, 2004; Matusz and Eimer, 2011; reviewed in De Meo et al., 2015; Murray et al. 2015). This notion has been challenged by other behavioural (e.g., Sanabria et al., 2007), event-related potential (ERP; e.g., Talsma, Doty and Woldorff, 2007; Matusz and Eimer, 2013), and neuroimaging (e.g., Fairhall and Macaluso, 2009) findings, supporting the claim that attention can affect multisensory outcomes (reviewed in van Atteveldt, Murray, Thut and Schroeder, 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 prior to 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 prior to 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,

1
2
3 where context can be understood as the immediate situation in which a stimulus is
4 presented (van Atteveldt et al., 2014). Until recently, in studies of multisensory processes
5 these contextual influences have been studied in terms of long-term experience and
6 learning. Studies of speech and communication signals provide clear examples of this (e.g.
7 Froyen et al., 2009; Matusz et al., 2015a). More recently, consideration has been given to
8 context-based influences operating on a shorter timescale. This can include both on-line
9 effects arising during the course of the experiment (e.g., Powers et al., 2009; Thelen et al.,
10 2014; Matusz et al., 2015b; von Kriegstein and Giraud, 2006) as well as effects at even finer
11 time-scales, such as those transpiring at an inter-trial level (Wylie et al., 2009; Murray et al.,
12 2009; King et al., 2012; Sandhu and Dyson, 2013). For example, the fact of either repeating a
13 task across two successive trials or switching from one task to another can dramatically
14 influence performance, such that reaction times and error rates both increase after
15 switching to perform a new task – i.e., the switch cost effect (reviewed in Wylie and Allport,
16 2000).

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36 The importance of top-down control for stimulus processing has been broadly studied in
37 cognitive control research using interference paradigms, where congruence between
38 different dimensions of two unisensory stimuli is manipulated (e.g., the flanker task;
39 Corballis and Gratton, 2003; Appelbaum et al., 2011). Real-world environments are typically
40 multisensory in nature and therefore congruence (or lack thereof) permeates it. For
41 example, the congruence of low-level spatio-temporal features has been identified as a
42 major determinant for the occurrence of multisensory integration (reviewed in Murray and
43 Wallace, 2012; Stein, 2012). By contrast, congruence has been largely under-investigated
44 with regard to the interplay of top-down control and multisensory processes.

45
46
47
48
49
50
51
52
53
54
55
56
57 Recently, Sarmiento et al. (2012) tested how irrelevant stimuli from one sense can
58
59
60

1
2
3 interfere with the processing of stimuli in another sense by creating distinctive contexts.
4
5 Participants discriminated the duration (short or long) of a visual stimulus that was
6
7 accompanied by a congruent or incongruent auditory distracter. Critically, Sarmiento et al.
8
9 (2012) manipulated also the likelihood of congruence between the duration of visual and
10
11 auditory stimuli (hereafter context of congruence) by means of spatial locations of the
12
13 auditory-visual stimuli. That is, in one spatial location (i.e., upper visual field) visual and
14
15 auditory stimuli would match in their duration on the majority of trials, in contrast to the
16
17 other location where on the majority of trials the stimuli would be mismatching. The
18
19 difference in performance across trials containing congruent versus incongruent auditory-
20
21 visual stimuli (hereafter congruence effect) was employed as an index of crossmodal
22
23 interactions. The manipulation of context of congruence thus served as a proxy for
24
25 understanding the importance of statistical learning for these multisensory processes (Altieri
26
27 et al., 2015; Baier et al., 2006; Barakat et al., 2013; Barenholtz et al., 2014; Beierholm et al.,
28
29 2009; Chandrasekaran et al., 2009).

30
31
32
33
34
35
36 Sarmiento et al. (2012) demonstrated that top-down attention did affect crossmodal
37
38 interactions; these were larger in contexts where there was a high versus low probability of
39
40 congruence. This was primarily due to a reduction in the congruence effect in the low-
41
42 probability context. The authors argued that this situation triggered an attentional set for
43
44 filtering out temporally incongruent auditory input that in turn resulted in larger crossmodal
45
46 interactions. It is likely that these context-based adjustments of top-down attentional
47
48 control are transpiring in an on-line fashion, as indicated by studies of purely visual
49
50 incongruence (King et al., 2012). If this holds true also for auditory-visual congruence-based
51
52 interactions, such influences should also operate on a trial-to-trial basis, which can be
53
54 investigated by separately quantifying crossmodal interactions (viz. congruence effects) on
55
56
57
58
59
60

1
2
3 trials when the context of congruence switches versus repeats. This was the principal aim of
4
5 the current study.
6

7
8 We achieved this by using the same paradigm as in Sarmiento et al., (2012) (hereafter,
9
10 visual temporal discrimination task) (**Figure 1**). We expected to replicate the behavioural
11
12 results of Sarmiento et al. (2012), by finding reliable influences of context on congruence-
13
14 based interactions, as tested using the visual temporal discrimination task. In the present
15
16 study, however, we expected that repetition of context versus a switch from one context to
17
18 another on successive trials would have opposing effects on this influence. Specifically, for
19
20 conditions where the context repeated we expected a larger congruence effect for the high-
21
22 probability versus low-probability context, which would not only replicate the findings of
23
24 Sarmiento et al. (2012) but also extend them to show that statistical learning about
25
26 contingencies between the stimulus context (i.e., a spatial location) and stimulus content
27
28 can take place across single trials. By contrast, for conditions where the context switched we
29
30 expected a larger congruence effect for the low-probability versus high-probability context,
31
32 because the attentional set cannot be reliably established.
33
34
35
36
37

38
39 By including electroencephalographic measurements analysed within an electrical
40
41 neuroimaging framework, we were likewise able to determine the spatio-temporal
42
43 correlates of context-based effects related to statistical learning as well as those related to
44
45 transitioning from one to another context from one trial to the next one. We expected to
46
47 differentiate between the possible contributions of modulations in brain response strength,
48
49 topography, and latency that could account for the influence of context on congruence-
50
51 based auditory-visual interactions. In this way, we were likewise able to situate the effects
52
53 found in the visual temporal discrimination task with respect to non-linear multisensory
54
55 integration processes as revealed by a control oddball task. In the latter, we expected to
56
57
58
59
60

1
2
3 observe non-linear response interactions between ERPs to AV stimuli and the summed
4
5 responses to unisensory stimuli (i.e., A + V) (e.g., Cappe et al., 2010, 2012).
6
7

8 **Methods**

9 *Participants*

10
11 The participants were 28 undergraduate students from the University of Granada (19
12
13 females; age range: 18-28 years old; mean age of 24 years) who received course credits or
14
15 cash payment of 20 Euro in exchange of their participation. All of the participants reported
16
17 normal hearing and normal or corrected-to-normal vision and gave their informed consent
18
19 to participate in the study conducted in accordance with the ethical standards laid down in
20
21 the 1964 Declaration of Helsinki. The data from two participants were removed due to
22
23 technical issues with Net Station software. Additionally, EEG data from six participants were
24
25 heavily contaminated by blink, muscle, or alpha-related artefacts, which resulted in less than
26
27 100 accepted EEG epochs for each condition of interest. Consequently, data from these
28
29 subjects were removed from analyses. The data from the remaining 20 participants (12
30
31 women; age range: 18-28 years, mean age = 23 years) were included in the complete data
32
33 analyses presented here.
34
35
36
37
38
39
40
41
42

43 *Apparatus and stimuli*

44
45 The experiment was conducted on an Intel Core 2 Duo PC with a 17-in. LCD monitor. E-
46
47 Prime software was used for stimulus presentation and response collection (Psychology
48
49 Software Tools; www.pstnet.com).
50
51

52
53 The visual stimuli consisted of a white circle (3.01° in diameter), a white equilateral
54
55 triangle (3.01° in height) and a white central cross that served as the fixation point. The
56
57 horizontal line of the cross (11.52° long) split the screen into upper and lower halves. Two
58
59
60

1
2
3 loudspeakers, positioned on each side of the computer screen, were used to present the
4
5 auditory stimuli, which consisted of a white noise burst and a 600 Hz pure tone (60 dB[A]
6
7 measured at ear level; see **Figure 1**).

9 10 ***Procedure***

11
12 Participants sat in a comfortable chair at approximately 57 cm from the computer
13
14 monitor in a silent, dark and electrically shielded room. The instructions that explained the
15
16 tasks were displayed on the computer screen. The experiment began with an unisensory
17
18 discrimination task in which participants were asked to discriminate the duration (short or
19
20 long) of a white circle displayed above or below (6.52° from the centre of the circle; cf.
21
22 Experiment 4 in Sarmiento et al., 2012) the fixation point (i.e., the white horizontal line) for
23
24 either 100 ms or 180 ms. The unisensory block (16 short and 16 long unisensory trials) was
25
26 used to ensure that participants understood the task and were responding above 60 % of
27
28 accuracy overall.
29
30
31
32

33
34 After the unisensory task participants put on an elastic electrode-cap for the EEG
35
36 recording (see EEG acquisition section below) during an oddball and a visual temporal
37
38 discrimination task. The oddball task was conducted first in order to avoid any carryover
39
40 effect from the context of congruence manipulation. However, as the visual temporal
41
42 discrimination task was the main focus of the present study, we describe it first here.
43
44
45
46
47

48 ***Visual temporal discrimination task***

49
50 The oddball task was followed by the visual duration discrimination task in which
51
52 participants were asked to discriminate the duration of a white circle displayed below or
53
54 above the fixation point for either 100 ms or 180 ms, while ignoring the synchronous
55
56 presentation of a white noise burst that could last for 100 ms or 180 ms. Onsets were always
57
58
59
60

1
2
3 synchronous across sensory modalities. No unisensory trials were presented in the visual
4
5 duration discrimination task. Congruent and incongruent trials emerged from the four
6
7 possible stimuli combinations. The crucial manipulation consisted of the inclusion of two
8
9 contexts of congruence defined by the two halves of the computer screen (upper/lower). For
10
11 half of the participants, the upper spatial location was associated with a high proportion of
12
13 congruent trials (80%) and the lower spatial location was associated with a low proportion of
14
15 congruent trials (20%). The reverse was true for the remaining half of the participants. The
16
17 congruent and incongruent stimuli had the same probability of appearance. This was also
18
19 true for short and long trials. Consequently, the context of congruence could repeat or
20
21 switch across trials with equal probability. Notably, stimuli with identical physical properties
22
23 were presented across the two contexts. This ensured that any differences observed
24
25 between these contexts could be attributed to stimulus congruence probabilities being
26
27 utilised by the brain to adjust top-down attentional control to facilitate visual performance
28
29 by suppressing the processing of sounds in contexts with a high proportion of auditory-visual
30
31 incongruence. This interpretation is in line with emerging notions regarding the interplay
32
33 between top-down attention mechanisms and stimulus expectations (e.g. Summerfield and
34
35 Egner, 2009; Kok et al. 2012; Larsson and Smith, 2012; Auksztulewicz and Friston, 2015).
36
37
38
39
40
41
42

43 At the beginning of each trial, participants were presented with the fixation cross
44
45 displayed in white against a black background for a random duration between 500 and 1000
46
47 ms. This fixation cross remained on throughout the whole trial. Next, the visual and auditory
48
49 stimuli were presented for either 100 ms or 180 ms, and could be congruent or incongruent
50
51 in duration. Feedback regarding response accuracy was provided to the participants for 500
52
53 ms. The next trial began between 1500 and 2000 ms after the feedback. Participants were
54
55 encouraged to fixate and to avoid eye movements and blinks during the presentation of the
56
57
58
59
60

1
2
3 stimuli. The sequence of events is schematised in **Figure 1a**.
4

5 Accuracy, rather than response speed, was emphasised; so participants had no response
6 time pressure. Responses were collected using a serial response box. Half of the participants
7 were told to press the leftmost key if the circle was short in duration and the rightmost key if
8 the circle was long in duration. The reverse stimulus-response mapping was used for the
9 remaining participants. All responses were made with the right hand. Participants completed
10 8 multisensory practice trials and 8 blocks of 120 multisensory trials each, distributed as
11 follows: 48 congruent trials and 12 incongruent trials for the high proportion-congruent
12 context; 12 congruent and 48 incongruent trials for the low proportion-congruent
13 context.
14
15
16
17
18
19
20
21
22
23

24 ***Oddball task***

25
26 By means of a control, oddball task, we likewise assessed whether the context-based
27 effects in the current experiment occur before or after any non-linear brain response
28 interactions between the stimuli we employed here. Importantly, we used the same stimuli
29 as in the visual temporal discrimination task, but made participants attend to stimulus
30 dimensions other than stimulus duration (now visual shape and sound frequency) while
31 removing the influence of context-based control processes (**Figure 1**). These manipulations
32 enabled us to verify whether there are qualitative changes in how the congruent and
33 incongruent stimuli from the visual temporal discrimination task are integrated when
34 presented outside of strong top-down influences based on task-relevance or context. A
35 corollary question here was whether congruence would further modulate these auditory-
36 visual interactions and, if so, at what stage. While outside of the primary focus of the current
37 study, this important issue has thus far received limited treatment (Fort et al., 2002;
38 Molholm et al., 2004; Yuval-Greenberg and Deouell, 2009).
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Participants performed an oddball task in which either a white triangle or a 600 Hz tone
4
5 had to be detected via button-press. The standard stimuli, which did not require a response,
6
7 were a white circle and a white noise burst (these were identical to those in the visual
8
9 temporal discrimination task described above). The visual stimuli could appear above or
10
11 below the fixation cross (6.52° from the centre of the circle). Participants performed two
12
13 blocks of 216 trials each, 11% of which were oddballs (i.e., 24 trials). Oddballs (triangles or
14
15 600Hz tones) and standards (circles or noise bursts) could last for 100 or 180 ms and could
16
17 be presented alone (unisensory stimuli) or synchronously with a stimulus of the other
18
19 sensory modality (multisensory stimuli) with simultaneous onsets. The visual and auditory
20
21 oddballs were never presented together. This manipulation gave rise to unisensory trials
22
23 (the visual or auditory stimulus was presented alone), multisensory congruent trials (the
24
25 visual stimulus was of the same duration as the auditory stimulus) and multisensory
26
27 incongruent trials (the visual stimulus was of a different duration than the auditory
28
29 stimulus). Unisensory and multisensory stimuli, short and long stimuli, as well as congruent
30
31 and incongruent stimuli had the same probability of presentation (**Figure 1b**). Participants
32
33 had to detect the oddballs when they appeared alone as well as when they were presented
34
35 accompanied by a standard stimulus from the other modality. The fixation point was
36
37 presented alone for 750 or 1250 ms, and the inter-trial interval ranged from 1500 to 2500
38
39 ms to avoid anticipation. Feedback was only displayed when participants made an error. The
40
41 oddball task was used to assess multisensory integration between congruent and
42
43 incongruent audiovisual stimuli when no response was required (i.e., standard stimuli).
44
45
46
47
48
49
50
51

52 At the end of the experiment, participants completed a questionnaire that evaluated
53
54 whether or not they were aware of the congruence manipulation in the visual temporal
55
56 discrimination task. They were asked whether they perceived any difference between the
57
58
59
60

1
2
3 two onset locations during the visual duration discrimination task, and, if so, what that
4
5 difference was.
6
7
8
9

10 **EEG Acquisition**

11
12
13 Continuous EEG was acquired at 1000 Hz through a Geodesic Sensor Net of 129
14
15 Ag/AgCl electrodes referenced to the vertex channel (Tucker et al., 1994) and connected to
16
17 an AC-coupled high-input impedance amplifier (200 M Ω , Net Amps™, Electrical Geodesics,
18
19 Eugene, Oregon). Impedances were kept below 50k Ω , as recommended for these amplifiers.
20
21 The signal was acquired with a 0.01-100Hz elliptical band-pass filter. Gain and zero
22
23 calibration were performed prior to the start of every recording. The head-coverage
24
25 included sensors lateral to and below both eyes to monitor horizontal and vertical eye
26
27 movements (electrooculogram, EOG).
28
29
30
31
32
33

34 **ERP Analyses**

35
36
37 The EEG was high-pass (0.1Hz) and low-pass (30Hz) filtered offline. Epochs of EEG were
38
39 segmented from 200 ms pre-stimulus to 800 ms post-stimulus onset. The segmented epochs
40
41 were submitted to automated software processing for identification of artefacts. Epochs
42
43 containing eye movements or blinks (± 70 μ V relative to baseline in EOG channels) were
44
45 rejected. Individual channels containing other sources of transient noise (± 80 μ V relative to
46
47 baseline in any channel) were replaced using a trial-by-trial basis with a spherical
48
49 interpolation algorithm (Perrin et al., 1989). The epoch was discarded when more than 10
50
51 channels were deemed artefact-contaminated. Additionally, the data were visually inspected
52
53
54
55
56 to ensure that all artefacts had been detected.

57
58 Epochs of EEG were averaged for each stimulus condition and from each subject to
59
60

1
2
3 calculate event-related potentials (ERPs). We would remind the reader that long and short
4
5 duration stimuli were constituents for both congruent and incongruent stimulus conditions.
6
7 By collapsing across durations, we were able to equate any contributions of offset responses
8
9 to between-condition effects as well as to ensure that the comparison of responses to
10
11 congruent and incongruent pairings involved equivalent stimulus energies. Data from ERPs
12
13 were baseline corrected using the period of 200 ms preceding stimulus onset. ERPs were
14
15 recalculated against the average reference.
16
17

18
19 For the visual temporal discrimination task, the averaging of EEG epochs was guided by
20
21 the pattern of behaviour. As will be detailed below, we observed a larger congruence effect
22
23 for high- than low- probability contexts when context repeats and a larger congruence effect
24
25 for low- than high- probability context when context switches (see **Figure 2**). Consequently,
26
27 we collapsed across trials with larger congruence effects as well as across trials with smaller
28
29 congruence effects. This enabled us to identify brain mechanisms of the congruence effect,
30
31 which serves as a proxy for audiovisual cross-modal interactions. In this way, there were 4
32
33 ERPs per participant following a 2x2 factorial design (Congruence effect size x Congruence):
34
35
36 1) L_{CONG} (i.e., collapsed trials when the duration was congruent and high-probability context
37
38 repeated as well as when the duration was congruent and low-probability context switched),
39
40
41 2) L_{INCONG} (i.e., collapsed trials when the duration was incongruent and high-probability
42
43 context repeated as well as when the duration was incongruent and low-probability context
44
45 switched), 3) S_{CONG} (i.e., collapsed trials when the duration was congruent and low-
46
47 probability context repeated as well as when the duration was congruent and high-
48
49 probability context switched), and 4) S_{INCONG} (i.e., collapsed trials when the duration was
50
51 incongruent and low-probability context repeated as well as when the duration was
52
53 incongruent and high-probability context switched). For each of these 4 conditions there
54
55
56
57
58
59
60

1
2
3 were on average (\pm s.e.m.) 184(\pm 9), 191(\pm 9), 191(\pm 9), and 182(\pm 9) accepted epochs,
4
5 respectively, with a minimum of 100 accepted epochs per subject and condition. Another
6
7 important advantage of collapsing the data in this manner is that any differences strictly due
8
9 to probability (and therefore to deviance detection) are counterbalanced across all
10
11 conditions. In this way, any differences between congruent and incongruent conditions
12
13 across contexts can be confidently interpreted purely in terms of top-down control.
14
15

16
17 For the oddball task, ERPs (collapsed across durations) were calculated for unisensory
18
19 visual trials (V), unisensory auditory trials (A), congruent auditory-visual trials (AV_{cong}), and
20
21 for incongruent auditory-visual trials (AV_{incong}). The ERPs from the A and V conditions were
22
23 summed (A+V) and then statistically compared with the AV_{cong} and AV_{incong} conditions in a 1-
24
25 way non-parametric F-test. Non-linear multisensory interactions were identified in planned
26
27 contrasts (AV_{cong} vs. A+V as well as AV_{incong} vs. A+V). Finally, planned contrasts tested for
28
29 effects of congruence (AV_{cong} vs. AV_{incong}). For the AV_{cong} , AV_{incong} , and A+V conditions there
30
31 were on average (\pm s.e.m.) 71(\pm 4), 71(\pm 3), and 72(\pm 4) accepted epochs, respectively, with a
32
33 minimum of 34 accepted epochs per subject and condition.
34
35
36

37
38 All analyses were conducted using the freeware Cartool
39
40 (<http://sites.google.com/site/fbmlab/cartool>; Brunet et al., 2011) and the STEN toolbox
41
42 (<http://www.unil.ch/line/home/menuinst/infrastructure/software--analysis-tools.html>). The
43
44 analysis strategy we used followed a multi-step analysis procedure referred to as electrical
45
46 neuroimaging (Murray et al., 2008). Electrical neuroimaging allowed us to identify effects
47
48 using both local and global measures of the electric field at the scalp. This procedure
49
50 distinguishes between effects following from modulations in the strength of responses of
51
52 statistically indistinguishable brain generators and alterations in the configuration of these
53
54 generators (viz. the topography of the electric field at the scalp), as well as latency shifts in
55
56
57
58
59
60

1
2
3 brain processes across experimental conditions (Michel et al., 2004, 2009; Murray et al.,
4
5 2008; Michel and Murray, 2012; Altieri et al., 2015; Stevenson et al., 2014). Additionally, we
6
7 used the local autoregressive average distributed linear inverse solution (LAURA; Grave de
8
9 Peralta Menendez et al., 2001, 2004) to visualise and statistically contrast the likely
10
11 underlying sources of effects identified in the preceding analysis steps.
12
13

14
15 Global modulations in the strength of the electric field at the scalp were quantified by
16
17 calculating the global field power (GFP; Lehmann and Skrandies, 1980) for each subject and
18
19 stimulus condition. This measure represents the spatial standard deviation of the electric
20
21 field at the scalp at a given time point, and constitutes a reference-independent measure of
22
23 the ERP amplitude (Murray et al., 2008; Koenig and Melie-Garcia, 2010). Stronger electric
24
25 fields result in larger GFP values. GFP values were baseline corrected in order to legitimately
26
27 compare between conditions with different numbers of trials. GFP modulations were
28
29 analysed using a millisecond-by-millisecond non-parametric analysis of variance (ANOVA)
30
31 and non-parametric planned contrasts when warranted using the Mann-Whitney test.
32
33 Analyses were performed in conjunction with a 15 contiguous data-point temporal criterion
34
35 for significant effects to correct for temporal auto-correlation (only those effects where p-
36
37 values met the statistical threshold ($p \leq 0.05$) for more than 15 contiguous time points (i.e.,
38
39 ~15 ms at 1000 Hz sampling) were considered reliable; inspired by Guthrie and Buchwald,
40
41 1991).
42
43
44
45
46
47

48
49 In order to identify stable periods of electric field topography (hereafter template
50
51 maps), the collective post-stimulus group-average ERPs were subjected to hierarchical
52
53 clustering. The optimal number of stable ERP clusters (i.e., the minimal number of maps that
54
55 accounts for the greatest variance of the dataset) was determined using a modified
56
57 Krzanowski-Lai criterion (Murray et al., 2008). The clustering makes no assumption on the
58
59
60

1
2
3 orthogonality of the derived topographic template maps (De Lucia et al., 2010a, 2010b;
4
5 Pourtois et al., 2008; Koenig et al., 2014). Template maps identified in the group-level ERP
6
7 were entered into a fitting procedure wherein each template map was compared with the
8
9 moment-by-moment scalp topography on the individual subjects' ERPs from each condition
10
11 and was labelled according to the one with which it best correlated spatially. This fitting
12
13 procedure allows determining the total amount of time a given template map was observed
14
15 for a given condition across subjects. Statistical analysis of these values was performed with
16
17 an ANOVA. The results of this analysis revealed whether a given ERP was more often
18
19 described by one template map vs. another, which implied different stable map
20
21 configurations across conditions at the same moments in time.
22
23
24
25
26
27

28 ***Source Estimations***

30
31 The intracranial generators underlying the surface electric field were estimated using
32
33 a distributed linear inverse solution based on a Local Auto-Regressive Average (LAURA)
34
35 model comprising biophysical laws as constraints (Grave de Peralta Menendez et al., 2001,
36
37 2004). LAURA uses a realistic head model with a solution space of 4024 nodes, selected from
38
39 a 6 x 6 x 6 mm grid equally distributed within the gray matter of the Montreal Neurological
40
41 Institute's (MNI's) average brain. This algorithm selects the source configuration that better
42
43 mimics the biophysical behaviour of electric vector fields, that is, the estimated activity at
44
45 one point depends on the activity at neighbouring points according to electromagnetic laws
46
47 (Grave de Peralta and Gonzales Andino, 2002). Prior to calculation of the inverse solution,
48
49 the ERP data of each individual subject for each condition were down-sampled and affine-
50
51 transformed to a common 111-channel montage. The time period for which intracranial
52
53 sources were estimated and statistically compared between conditions was defined by the
54
55
56
57
58
59
60

1
2
3 results of the abovementioned topographic analysis. Statistical analyses of source
4
5 estimations were performed by first averaging the ERP data across time to generate a single
6
7 data point for each participant and condition. ANOVAs were performed at each of the 4024
8
9 source nodes in the inverse solution space (using across-subjects variance). For the temporal
10
11 discrimination task, the ANOVA followed a 2x2 factorial design (congruence x congruence
12
13 effect size). For the oddball task this ANOVA followed a 1-way design (AVcong, AVincong,
14
15 A+V). Only nodes exceeding the spatial extent criterion of at least 17 contiguous significant
16
17 nodes ($p \leq 0.0002$) were considered significant (see Thelen et al, 2012 for similar spatial
18
19 criterion). This spatial criterion was determined using the AlphaSim program
20
21 (<http://afni.nimh.nih.gov>) and assuming a spatial smoothing of 6mm full-width half
22
23 maximum. AlphaSim performs 10,000 Monte Carlo permutations on the 4024 nodes of our
24
25 lead field matrix to determine the false discovery rate for clusters of different sizes. These
26
27 permutations indicated that there is a 3.54% probability of a cluster of at least 17 contiguous
28
29 nodes, which gives an equivalent node-level p-value of $p \leq 0.0002$. The results of the analyses
30
31 were rendered on the MNI average brain with loci of the largest statistical differences within
32
33 a cluster indicated based on the Talairach and Tournoux (1988) atlas.
34
35
36
37
38
39
40
41
42

43 **Results**

44 ***Visual Temporal Discrimination Task***

45 *Behavioural data*

46
47
48
49
50
51 Mean response accuracy for each participant and condition were submitted to a 2
52
53 Context of congruence (high congruent, low congruent) x 2 Congruence (congruent,
54
55 incongruent) x 2 Context Transition (switch, repeat) parametric ANOVA. The first trial of
56
57 each block was excluded from the analysis. Neither the main effect of context transition
58
59
60

1
2
3 (F_(1,19)=1.91; $p>0.18$; $\eta_p^2=0.09$) nor context of congruence (F_(1,19)=0.153; $p=0.7$; $\eta_p^2<0.01$)
4
5 were reliable. There was a reliable main effect of congruence (F_(1,19)=42.55; $p<0.001$; $\eta_p^2=$
6
7 0.69). None of the 2-way interactions were reliable (all F 's<1.69; p 's>0.20). Crucially, there
8
9 was a reliable 3-way interaction (F_(1,19)=18.66; $p<0.001$; $\eta_p^2=0.49$).
10
11

12 Additional 2-way ANOVAs for each level of the factor context transition were conducted
13
14 to understand the bases for this 3-way interaction. For the repeat context transition 2x2
15
16 ANOVA there was a reliable effect of congruence (F_(1,19)=36.31; $p<0.001$; $\eta_p^2=0.66$) as well
17
18 as a reliable interaction between context and congruence (F_(1,19)=11.67; $p<0.003$; $\eta_p^2=$
19
20 0.38). There was no main effect of context (F_(1,19)=0.36; $p>0.55$; $\eta_p^2=0.02$). The interaction
21
22 on trials involving repetition of context was driven by larger in magnitude congruence
23
24 effects in the high proportion-congruent context (23.8%) than in the low proportion-
25
26 congruent context (16.6%); both being significant (t 's>4.6; p 's<0.001) (**Figure 2**). For the
27
28 switch context transition 2x2 ANOVA there was likewise a reliable effect of congruence
29
30 (F_(1,19)=43.06; $p<0.001$; $\eta_p^2=0.69$) as well as a reliable interaction between context and
31
32 congruence (F_(1,19)=7.06; $p=0.016$; $\eta_p^2=0.27$). There was no main effect of context (F_(1,19)
33
34 =1.74; $p>0.20$; $\eta_p^2=0.08$). In contrast to the context-repeat trials, the interaction on trials
35
36 involving a switch in context between trials was driven by larger in magnitude congruence
37
38 effects in the low proportion-congruent context (23.3%) than in the high proportion-
39
40 congruent context (17.9%); both being significant (t 's>5.2; p 's<0.001). That is, there was a
41
42 larger congruence effect when switching to a low proportion-congruent context.
43
44
45
46
47
48
49

50 Participants' reports revealed that five (25%) of the participants noticed a difference
51
52 between the two contexts (e.g., "I made more errors on the upper half"). However, none of
53
54 them reported being aware of the specific proportion-congruent manipulation, in
55
56 accordance with previous studies (Crump et al., 2006; Heinemann et al., 2009; Sarmiento et
57
58
59
60

1
2
3 al., 2012; King et al., 2012).
4
5
6

7
8 Surface ERP data
9

10 In light of the behavioural results described above, the ERP data were analysed
11 following a 2x2 Congruence effect size x Congruence factorial design (see Materials and
12 Methods for details). The timepoint-wise non-parametric ANOVA on the Global Field Power
13 waveforms revealed a significant interaction over the 228-257 ms period (**Figure 3**). Non-
14 parametric post-hoc tests within this period showed there to be a significantly weaker GFP
15 for L_{CONG} than L_{INCONG} over the 240-259ms period. By contrast, there was a significantly
16 stronger GFP for S_{CONG} than S_{INCONG} over the 224-242ms period. Neither main effect yielded
17 significant results.
18
19
20
21
22
23
24
25
26
27

28 A hierarchical topographic cluster analysis was performed on the collective group-
29 average ERPs to identify periods of stable electric field topography both within and between
30 experimental conditions. 11 template maps that explained 95.9% of the variance were
31 identified. At the group-average level there was no evidence for different template maps
32 across conditions (see **Supplementary Figure 1**). Consequently, no single-subject fitting was
33 performed with these data.
34
35
36
37
38
39
40
41
42
43
44

45 Source estimations
46

47 Distributed source estimations were calculated using the mean potential values over
48 the 228-257 ms period and submitted to a 2 Congruence \times 2 Congruence-effect size ANOVA.
49 The main effect of congruence yielded no significant clusters. The main effect of congruence
50 effect size yielded significant clusters centred within the left middle occipital gyrus, the right
51 parahippocampal gyrus, the right precuneus, and the right inferior frontal gyrus. These
52
53
54
55
56
57
58
59
60

1
2
3 regions are consistent with the literature reporting their contribution to cross-modal and
4
5 multisensory processes (reviewed in Murray and Wallace, 2012), including those focusing on
6
7 stimulus (a)synchrony (e.g., Macaluso et al., 2004). A significant interaction between
8
9 congruence and congruence-effect size was observed in a widely distributed network of
10
11 clusters within the left superior temporal cortices, the left parietal cortices, and bilateral
12
13 precuneus (**Figure 4a**). In particular, the parietal cortices likely included pre-motor structures
14
15 previously implicated in temporal processes (e.g., Grahn and Rowe, 2009; Marchant et al.,
16
17 2012). Given this interaction, contrasts focused on the congruence effect for the large-size
18
19 congruence effect (L_{CONG} vs. L_{INCONG}) and small-size congruence effect (S_{CONG} vs. S_{INCONG})
20
21 conditions, separately. For the large-size congruence effect conditions, clusters exhibiting
22
23 significant differences were centred within the left medial frontal gyrus, the left superior
24
25 temporal cortex, and the precuneus bilaterally (**Figure 4b**). All clusters were more strongly
26
27 active in response to incongruent than congruent trials. For the small-size congruence effect
28
29 conditions, clusters exhibiting stronger responses to congruent than incongruent pairings
30
31 were centred within the left parietal cortices (**Figure 4c**).
32
33
34
35
36
37
38
39

40 **Oddball Task**

41 Behavioural data

42
43 Overall accuracy across participants on the oddball task was 99.77%, with a mean
44
45 reaction time of 340 ms. This high performance provides an assurance that participants were
46
47 appropriately engaged in the oddball task.
48
49

50 Surface ERP data

51
52 Analyses of data from the Oddball task focused on determining differences in the neural
53
54 activation between the ERPs in response to the congruent and incongruent multisensory
55
56
57
58
59
60

1
2
3 conditions and the summed ERPs in response to the unisensory auditory and visual
4
5 conditions (denoted by AV_{cong} , AV_{incong} and A+V, respectively). Only trials for standard stimuli
6
7 were included in the analyses, because they constituted the pool of stimuli used for the
8
9 visual temporal discrimination task, the analyses of which are described above.
10
11

12 The GFP waveforms were compared statistically using a 1-way ANOVA as a function of
13
14 time (**Figure 5**). There was a significant main effect over two time intervals (130-145 ms and
15
16 249-800 ms). Follow-up planned nonparametric contrasts revealed no significant differences
17
18 between AV_{cong} and AV_{incong} responses. Likewise, for both time windows, there were super-
19
20 additive multisensory responses that were observed irrespective of stimulus congruence.
21
22

23
24 Next, a hierarchical spatio-temporal cluster analysis was employed again to test for
25
26 topographic differences between experimental conditions within the data from the oddball
27
28 task. This procedure identified a set of 11 template maps explaining 96.3% of the variance
29
30 of the concatenated group-averaged ERP data set. There was no evidence for different
31
32 template maps across conditions (see **Supplementary Figure 2**). Consequently, no single-
33
34 subject fitting was performed with these data.
35
36
37
38
39

40 Source estimations

41
42
43 Estimated intracranial sources over the 130-145 ms post-stimulus time period
44
45 revealed a main effect of stimulus condition centred within a distributed network of areas
46
47 including the calcarine sulcus bilaterally, the left parieto-occipital cortex and left parietal
48
49 cortices, the orbitofrontal cortex bilaterally and the right lateral occipital cortices (see
50
51 **Figure 6a**). These areas have been repeatedly observed in studies of audiovisual integration
52
53 (e.g., Calvert et al., 2000, Dhamala et al., 2007; Cappe et al., 2010, 2012), with the
54
55 orbitofrontal cortices implicated in multisensory conflict (Diaconescu et al., 2011). Planned
56
57
58
59
60

1
2
3 comparisons between the AV_{cong} and A+V conditions revealed super-additive responses
4
5 within the calcarine sulcus bilaterally, the left parieto-occipital cortex and left parietal
6
7 cortices, and sub-additive responses within the orbitofrontal cortex bilaterally (**Figure 6b**). A
8
9 similar pattern of effects was observed between AV_{incong} and A+V, with the addition of
10
11 super-additive responses within the right lateral occipital cortices (**Figure 6c**). Finally,
12
13 contrasts between AV_{cong} and AV_{incong} revealed stronger responses to the AV_{cong} condition
14
15 within the calcarine sulcus bilaterally and stronger responses to the AV_{incong} condition within
16
17 the right lateral occipital cortex, the right parietal cortex and the bilateral orbitofrontal
18
19 cortex, the right parietal cortex and the bilateral orbitofrontal
20
21 cortex (**Figure 6d**). However, we are hesitant to over-interpret these results, given that
22
23 corresponding effects were not reliably at the level of scalp ERPs.
24
25

26
27 Source estimations over the 195-275 ms post-stimulus time period revealed a main
28
29 effect of stimulus condition throughout an extensive brain network including bilateral
30
31 occipital, parietal, temporal, and frontal sources (**Supplementary Figure 3**). Planned
32
33 comparisons between either AV_{cong} and A+V or AV_{incong} and A+V revealed sub-additive
34
35 responses within this network. The contrast between AV_{cong} and AV_{incong} revealed stronger
36
37 responses to AV_{cong} within the left temporal pole, the anterior cingulate, and dorsal
38
39 prefrontal cortex. Stronger responses to the AV_{incong} condition were observed in right
40
41 parietal cortices. This locus is similar to that observed for the large-size congruence effect
42
43 contrast (see **Figure 4b**).
44
45
46
47
48
49

50 Discussion

51
52
53 It is well established that some multisensory processes can occur in a bottom-up
54
55 fashion, i.e., independently of top-down control based on one's goals. Much less is known
56
57 about how these processes are controlled by influences based on the context in which the
58
59
60

1
2
3 multisensory stimuli are presented as well as whether (and if so how) context-based
4
5 mechanisms operating across different timescales interact with each other.
6

7
8 In a series of behavioural experiments, Sarmiento et al. (2012) demonstrated that the
9
10 ability of an irrelevant auditory stimulus to interfere with a duration judgement on a visual
11
12 stimulus can be influenced by the particular context in which the stimuli appear. Specifically,
13
14 spatial locations where there was a large proportion of congruent auditory-visual stimuli
15
16 were generally giving rise to stronger influences of the sound on visual processing than in
17
18 spatial locations with higher proportion of incongruent auditory-visual presentations. The
19
20 authors interpreted these findings in terms of specific contexts activating top-down
21
22 attentional control mechanisms to a different extent. Contexts with a high proportion of
23
24 incongruent pairings trigger enhanced top-down inhibition that in turn attenuated auditory-
25
26 visual interactions, indexed by behavioural congruence effects. This top-down control
27
28 enhancement likely occurred implicitly, as indicated by the majority of participants being
29
30 unaware of the experimental manipulation.
31
32
33
34
35

36 The present study employed the same paradigm to understand whether context-based
37
38 control operates at a finer 'online' time scale. To do this, we analysed trials when the
39
40 context of congruence repeated versus switched on two successive trials, and
41
42 electrophysiological recordings were used to understand the brain mechanisms underlying
43
44 the influence of these control mechanisms. Furthermore, we added a control, oddball task,
45
46 where the same stimuli as those used on the visual temporal discrimination task were both
47
48 made irrelevant, so that the timing of the non-linear interactions between these stimuli
49
50 could be assessed outside of influences of top-down control and then compared with the
51
52 timing of the occurrence of context-based influences described above.
53
54
55

56
57 The visual temporal discrimination task revealed that congruence effects differed in
58
59
60

1
2
3 magnitude depending on the context of congruence as well as on whether a particular
4
5 context repeated or switched. Specifically, larger-size congruence effects were observed for
6
7 high-proportion congruence contexts on trials where the context repeated and for low-
8
9 proportion congruence contexts on trials where the context switched (Figure 2). Our
10
11 analyses of the brain responses focused on the mechanisms giving rise to large versus small
12
13 congruence effects at the level of behaviour. Audiovisual interactions were modulated by
14
15 context transition at ~230ms post-stimulus, with these modulations being driven by changes
16
17 in brain response strength within a statistically indistinguishable configuration of brain areas
18
19 (Figure 3). In the case of large-size congruence effects, brain responses were attenuated on
20
21 trials with congruent than incongruent auditory-visual stimuli, and originating in the left
22
23 medial frontal gyrus, the left superior temporal cortex, and the precuneus bilaterally (Figure
24
25 4). Small-size congruence effects were found to be associated with stronger responses on
26
27 congruent than incongruent trials in the left parietal cortices.
28
29
30
31
32

33
34 Auditory-visual interactions in the visual temporal discrimination task differed in
35
36 their strength depending on which context they appeared in as well as whether this context
37
38 repeated or switched across two successive trials. Larger (and comparable in size)
39
40 congruence effects were observed in contexts with high proportion of auditory-visual
41
42 congruence on trials where this context repeated and in contexts with high proportion of
43
44 cross-modal *incongruence* when this context *switched*. These results can be most readily
45
46 interpreted in terms of the stimulus context (or rather contexts operating both at coarser
47
48 and finer time scales) modulating the cross-modal interactions by means of a form of
49
50 statistical learning.
51
52
53

54
55 Statistical learning is typically understood as a process by which a cognitive system
56
57 learns about the underlying structure of the sensory inputs by extracting the distributional
58
59
60

1
2
3 properties from these inputs across time and/or space (Saffran et al., 1996). While the
4
5 domain-general versus -specific nature of this process is a topic of current debate (Frost et
6
7 al., 2015), it has been demonstrated to support a variety of cognitive functions, from
8
9 auditory stimulus parsing and visual search to conditioning, to name just a few (e.g., Saffran
10
11 et al., 1996; Baker et al., 2004; Courville et al., 2006; Goujon and Fagot, 2013). The
12
13 importance of statistical learning has likewise been demonstrated for multisensory
14
15 processing (Baier et al., 2006; Beierholm et al., 2009; Chandrasekaran et al., 2009; Barakat
16
17 et al., 2013; Barenholtz et al., 2014; Altieri et al., 2015). For example, one can learn to
18
19 associate arbitrary but spatially and temporally congruent auditory and visual stimuli (pure
20
21 tones and Gabor patches, respectively), with reduced brain response strength corroborating
22
23 increased efficiency of their recognition after a few days of training (Altieri et al., 2015). In a
24
25 similar vein, learning of pairs consisting of human voices and gender-congruent faces is
26
27 facilitated compared to pairs involving human voices and images of gender-incongruent
28
29 faces or of plants/ rocks (Barenholtz et al., 2014).
30
31
32
33
34
35

36 How do the current results fit in with our understanding of the role of statistical
37
38 learning and context in controlling multisensory processing? Top-down attentional control
39
40 would be a prime candidate for a mechanism that could to link a particular location, a
41
42 particular audiovisual stimulus (congruent or incongruent) and, crucially, a particular context
43
44 of congruence (Treisman, 2006 and 1996). Indeed, research on implicit learning has
45
46 repeatedly shown that selective attention is critical in order to establish a link between
47
48 predictive dimensions of a particular stimulus (probability of congruence and spatial location
49
50 in this case), given that such associations only emerge when these dimensions are relevant
51
52 to the observer (Jiménez and Méndez, 1999; Crump et al., 2008; see also Kok et al., 2012;
53
54 Larsson and Smith, 2012; Auksztulewicz and Friston, 2015 for evidence on interactions
55
56
57
58
59
60

1
2
3 between top-down attention and expectations). In the present study, the location was
4
5 crucial since the participant needed to localize the target (displayed in one of two possible
6
7 locations) in order to perform the task. Once the association between proportion of
8
9 congruence and location was established, top-down attentional control of differing strength
10
11 was applied to perform the task within the particular context (see e.g. Vossel et al., 2014 for
12
13 a recent study modelling saccadic response-times data from a spatial-attention task). As a
14
15 result, the very same stimulus (e.g., congruent) was processed differently as a function of
16
17 the location where it was presented (that determined the context of congruence) and
18
19 whether this context had repeated or switched. An additional contributing factor may
20
21 therefore be whether or not the stimulus itself was repeated or switched (in addition to
22
23 whether the context itself repeated or switched). Unfortunately, in the present study, it was
24
25 infeasible to include this additional factor, principally because it would have severely
26
27 reduced the signal-to-noise ratio of the ERP data. Notwithstanding, stimulus repetition
28
29 effects in the current paradigm would be a particularly interesting avenue to pursue in
30
31 future research, given recent evidence for the interplay between stimulus repetition and
32
33 factors such as expectancy, attention, and memory (e.g., Henson et al., 2000; Turk-Browne
34
35 et al., 2007; Doehrmann et al., 2010; Recasens et al. 2014; reviewed in Segaert et al., 2013).

36
37
38 In a functional neuroimaging study employing a visual face-gender classification
39
40 version of the flanker task involving a location-based contextual manipulation of conflict
41
42 frequency, King et al. (2012) demonstrated that reduced interference in contexts with high
43
44 proportion of conflict is associated with stronger engagement of top-down control areas
45
46 (predominantly the medial superior parietal lobule), with their comparatively weak
47
48 engagement in contexts with a low proportion of conflict. The current results are in line with
49
50 differing engagement of top-down attentional control across the two contexts of
51
52
53
54
55
56
57
58
59
60

1
2
3 congruence. If a particular location becomes associated with a high proportion of
4
5 congruence between the visual and auditory stimuli, there is no need for heightened top-
6
7 down control as the presence of the irrelevant sounds would typically improve the visual
8
9 duration discrimination. The interactions arising between the auditory and visual stimuli are
10
11 likely to be larger in the case of a repeated context that involves weak top-down inhibition
12
13 of the irrelevant sounds. Along the same lines, a repeated context involving stronger top-
14
15 down inhibition, as in the case of the context involving a low proportion of cross-modal
16
17 congruence, is bound to lead to attenuated auditory-visual interactions when compared
18
19 with a switch within such a context.
20
21
22
23

24 Several important aspects of the current results support the interpretation in terms
25
26 of online and implicit adjustment of top-down attentional control based on a particular
27
28 context of congruence as well as whether it repeated or switched. First, we hasten to remind
29
30 the reader that the stimuli appearing in the two contexts were identical in their physical
31
32 properties (i.e., same number of auditory-visual congruent and incongruent trials). This
33
34 excludes the possibility that differences in brain responses were due to differences in
35
36 stimulation or stimulus-response associations for each context. Second, by collapsing the
37
38 ERP across conditions involving high-probability and low-probability contexts we also have
39
40 minimised any confounding influences on the brain activity stemming strictly from detection
41
42 of improbable events (e.g., P300, whose latency would coincide with the latency of our
43
44 effects; reviewed in Polich, 2007). Thus, the observed ERP differences can reasonably be
45
46 interpreted in terms of the brain utilising the probabilities with which stimuli were
47
48 congruent in particular contexts to adjust top-down attentional control. Lastly, the lack of
49
50 explicit awareness in the majority of participants regarding the congruence manipulation in
51
52 the visual temporal discrimination task accords with previous findings using similar
53
54
55
56
57
58
59
60

1
2
3 paradigms (e.g., Crump et al, 2008; Heinemann et al., 2009; King et al., 2012; Sarmiento et
4
5 al., 2012). This result suggests that the top-down attention adjustments were likely occurring
6
7 outside of the voluntary control of participants. This notion is further reinforced by the
8
9 differences observed between repeated and switched contexts, indicating the online (i.e.,
10
11 trial-by-trial) nature of these implicit top-down attention adjustments.
12
13

14
15 The current study demonstrates how important for cross-modal interactions is the
16
17 context in which multisensory stimuli appear, with its effects multiplexing across finer and
18
19 broader time scales to influence these interactions. A predominance of congruence between
20
21 signals across different senses will determine the level of top-down inhibition the brain
22
23 associates with a particular context. This will have a cascading effect, with stronger cross-
24
25 modal interactions transpiring across contexts associated with both high and low proportion
26
27 of auditory-visual congruence, on trial-to-trial variations in the magnitude of cross-modal
28
29 interactions (Murray et al., 2009; Shandu and Dyson, 2013). Importantly, the current results
30
31 point to the fact that incongruity does not necessarily trigger heightened top-down control.
32
33 More broadly, the current study sheds new light on the ongoing debate regarding the
34
35 influence of top-down control on multisensory processes (Talsma et al., 2010; van Atteveldt
36
37 et al., 2014; De Meo et al., 2015; Murray et al., 2015; ten Oever et al., in revision).
38
39
40
41
42

43 The employment of a control, oddball task enabled us to compare the timing of non-
44
45 linear multisensory interactions with that of cross-modal interactions elicited by the same
46
47 stimuli when under top-down control. Non-linear multisensory interactions first occurred in
48
49 the present dataset at 130-145ms post-stimulus, independent of the stimulus congruence.
50
51 By contrast, effects in the visual temporal discrimination task occurred first at ~230ms post-
52
53 stimulus (i.e. some 100ms later). While the bases for this latency difference remain to be
54
55 fully identified, these results corroborate an emerging consensus that timing of multisensory
56
57
58
59
60

1
2
3 interactions at the brain level might be one way of distinguishing processes less and more
4 dependent on top-down control (reviewed in De Meo et al., 2015). Some multisensory
5 processes, such as those based on simultaneity, seem to occur independently of top-down
6 attention and context influences, being observed across paradigms and task demands, as
7 well as across different species. These bottom-up, early multisensory interactions (eMSI) are
8 typically observed within 100ms post-stimulus onset (De Meo et al., 2015). Some of the
9 strongest support for their independence of top-down attention or context is provided by
10 studies reporting eMSI in anaesthetised preparations (reviewed in reviews in Sarko et al.,
11 2012; Rowland and Stein, 2014). In contrast, multisensory processes based on congruence
12 (e.g., perceptual as in the current study; see also Fort et al., 2002; semantic in Molholm et
13 al., 2004; Yuval-Greenberg and Deouell, 2009) typically transpire at later post-stimulus
14 stages. In the case of the present study, the timing differences across the two tasks may also
15 be linked to the fact that the visual temporal discrimination can at the earliest be completed
16 only after 100ms post-stimulus (i.e. at the duration of the shorter stimulus). Future work will
17 need to either modify the task to allow for its potential completion upon stimulus onset or
18 otherwise manipulate parametrically the said timing to more fully dissociate multisensory
19 effects differently dependent on top-down attention. Collectively, the documentation of
20 multisensory effects across multiple time scales corroborate a novel framework
21 differentiating top-down control processes based on attention (i.e., goals), context and
22 content (ten Oever et al., in revision). The current results support this proposal by showing
23 that both congruent and incongruent stimuli triggered non-linear multisensory response
24 interactions at ~140ms post-stimulus; in some instances stimulus congruence might not play
25 a role in modulating multisensory processing unless it is relevant to the current task (see also
26 Molholm et al., 2004, for similar multisensory findings; see Matusz et al., in revision, for
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 cross-modal results on predictability enabling the suppression of task-irrelevant sounds).
4
5
6

7 8 **Conclusions** 9

10
11 In a visual temporal discrimination task, stronger influences of irrelevant sound were
12
13 observed in repeated contexts associated with a high proportion of auditory-visual
14
15 congruence and following switch in context associated with a low proportion of congruence.
16
17 Multisensory processes are therefore subject to control both by context as well as by
18
19 statistical learning, where the latter operates on at least two time scales, but only in
20
21 situations where congruence across the senses is relevant for task demands. These findings
22
23 provide important insights into the interactions between expectations and top-down
24
25 attention occurring in environments closely resembling naturalistic ones, i.e. where stimuli
26
27 differ in their relevance to the current goals of the observer as well as the sensory modality
28
29 in which they appear.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure Captions

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.

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 (s.e.m. 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 $p < 0.001$).

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 2-way interaction was observed (228-257 ms). The inset displays an enlarged view of the 200-300 ms post-stimulus period.

Figure 4. The results of statistical analyses of distributed source estimations from the visual

1
2
3 temporal discrimination task are displayed. **a.** The 2x2 interaction yielded significant clusters
4
5 ($p < 0.05$; $kE > 17$ contiguous nodes) centred within the left middle occipital gyrus, the right
6
7 parahippocampal gyrus, the right precuneus, and the right inferior frontal gyrus. **b.** Follow-
8
9 up planned contrasts for the large-size congruence-effect revealed significant clusters
10
11 centred within the left medial frontal gyrus, the left superior temporal cortex, and the
12
13 precuneus bilaterally. All clusters were more strongly active in response to incongruent than
14
15 congruent trials. **c.** Follow-up planned contrasts for the small-size congruence-effect
16
17 revealed significant clusters centred within the left parietal cortices with stronger responses
18
19 to congruence than incongruent pairings.
20
21
22

23
24 **Figure 5.** Group-averaged global field power waveforms from the oddball task are displayed
25
26 for the AV_{cong} , AV_{incong} , and A+V conditions as a function of time. The red area plot displays
27
28 the time periods when a significant 1-way interaction was observed (130-145 and 249-800
29
30 ms; truncated in the figure at 500 ms).
31
32

33
34 **Figure 6.** The results of statistical analyses of distributed source estimations over the 130-
35
36 145 ms time period from the oddball task are displayed. **a.** The 1-way ANOVA yielded
37
38 significant clusters ($p < 0.05$; $kE > 17$ contiguous nodes) centred within a distributed network of
39
40 areas including the calcarine sulcus bilaterally, the left parieto-occipital cortex and left
41
42 parietal cortices, the orbitofrontal cortex bilaterally and the right lateral occipital cortices. **b.**
43
44 Follow-up planned contrasts between AV_{cong} and A+V revealed super-additive responses
45
46 within the calcarine sulcus bilaterally, the left parieto-occipital cortex and left parietal
47
48 cortices, and sub-additive responses within the orbitofrontal cortex bilaterally. **c.** A similar
49
50 pattern of effects was observed between AV_{incong} and A+V, with the addition of super-
51
52 additive responses within the right lateral occipital cortices. **d.** The contrast between AV_{cong}
53
54 and AV_{incong} revealed stronger responses to the AV_{cong} condition within the calcarine sulcus
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

bilaterally and stronger responses to the AV_{incong} condition within the right lateral occipital cortex, the right parietal cortex and the orbitofrontal cortex (bilaterally).

References

- 1
2
3
4
5
6 Altieri, N., Stevenson, R.A., Wallace, M.T. and Wenger, M. J. (2015). Learning to associate
7
8 auditory and visual stimuli: Behavioral and neural mechanisms. *Brain Topogr*, 28(3).
9
10 479-493. doi: 10.1007/s10548-013-0333-7.
11
12
13 Appelbaum, L.G., Smith, D.V., Boehler, C.N., Chen, W.D., and Woldorff, M.G. (2011). Rapid
14
15 modulation of sensory processing induced by stimulus conflict. *J Cogn Neurosci* 23(9),
16
17 2620-2628. doi: 10.1162/jocn.2010.21575
18
19
20 Auksztulewicz, R. and Friston, K. (2015). Attentional enhancement of auditory mismatch
21
22 responses: A DCM/MEG Study. *Cereb Cortex* 1-11. doi: 10.1093/cercor/bhu323
23
24
25 Baier, B., Kleinschmidt, A., and Müller, N. G. (2006). Cross-modal processing in early visual
26
27 and auditory cortices depends on expected statistical relationship of multisensory
28
29 information. *J Neurosci*, 26(47), 12260-12265. doi: 10.1523/JNEUROSCI.1457-06.2006
30
31
32 Baker, C.I., Olson, C.R., and Behrmann, M. (2004). Role of attention and perceptual grouping
33
34 in visual statistical learning. *Psyc Sci*, 15(7), 460-466. doi: 10.1111/j.0956-
35
36 7976.2004.00702.x
37
38
39 Barakat, B.K., Seitz, A.R., Shams, L. (2013). The effect of statistical learning on internal
40
41 stimulus representations: Predictable items are enhanced even when not
42
43 predicted. *Cognition*, 129, 205–211. doi:10.1016/j.cognition.2013.07.003
44
45
46 Barenholtz, E., Lewkowicz, D.J., Davidson, M., Mavica, L. (2014). Categorical congruence
47
48 facilitates multisensory associative learning. *Psyc Bull Rev*, 21, 1346–1352. doi:
49
50 10.3758/s13423-014-0612-7
51
52
53 Beierholm, U.R., Quartz, S.R., Shams, L. (2009). Bayesian priors are encoded independently
54
55 from likelihoods in human multisensory perception. *J Vision*, 9, 23. doi:
56
57 10.1167/9.5.23
58
59
60

- 1
2
3 Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.C. and Samson, Y.
4
5 (1998) Lateralization of speech and auditory temporal processing. *J Cogn Neurosci*,
6
7 10, 536–540. doi: 10.1162/089892998562834
8
9
10 Bertelson, P., and de Gelder, B. (2004). The psychology of multimodal perception. In C.
11
12 Spence and J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 141-
13
14 177). Oxford: Oxford University Press.
15
16
17 Brunet, D., Murray, M.M. and Michel, C.M. (2011). Spatiotemporal analysis of multichannel
18
19 EEG: CARTOOL. *Comput Intell Neurosci* 2011, 813-870. doi: 10.1155/2011/813870
20
21
22 Calvert, G.A., Campbell, R., and Brammer, M.J. (2000). Evidence from functional magnetic
23
24 resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr*
25
26 *Biol* 10, 649–657. doi: 10.1016/S0960-9822(00)00513-3
27
28
29 Cappe, C., Thut, G., Romei, V. and Murray, M.M. (2010). Auditory-visual multisensory
30
31 interactions in humans: timing, topography, directionality, and sources. *J Neurosci*
32
33 30(38), 12572–12580. doi: 10.1523/JNEUROSCI.1099-10.2010
34
35
36 Cappe, C., Thelen, A., Romei, V., Thut, G. and Murray, M.M. (2012). Looming signals reveal
37
38 synergistic principle of multisensory integration. *J Neurosci* 32, 1171–1182. doi:
39
40 10.1523/jneurosci.5517-11.2012
41
42
43 Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., Ghazanfar, A.A. (2009). The
44
45 natural statistics of audiovisual speech. *PLoS Comput Biol*, 5, e1000436. doi:
46
47 10.1371/journal.pcbi.1000436
48
49
50 Corballis, P.M. and Gratton, G. (2003). Independent control of processing strategies for
51
52 different locations in the visual field. *Biol Psychol* 64, 191-209. doi: 10.1016/S0301-
53
54 0511(03)00109-1
55
56
57 Courville, A.C., Daw, N.D., and Touretzky, D.S. (2006). Bayesian theories of conditioning in a
58
59
60

- 1
2
3 changing world. *Trends Cogn Sci*, 10(7), 294-300. doi:10.1016/j.tics.2006.05.004
4
5 Crump, M.J., Gong, Z. and Milliken, B. (2006). The context-specific proportion congruent
6
7 stroop effect: location as a contextual cue. *Psychon Bull Rev*, 13, 316-321.
8
9 doi:10.3758/ BF03193850
10
11 Crump, M. J. C., Vaquero, J. M. M., and Milliken, B. (2008). Context- specific learning and
12
13 control: The role of awareness, task relevance, and relative salience. *Conscious Cogn*,
14
15 17, 22–36. doi:10.1016/j.concog.2007.01.004
16
17
18
19 De Lucia, M., Clarke, S. and Murray, M.M., (2010a). A temporal hierarchy for conspecific
20
21 vocalization discrimination in humans. *J Neurosci*, 30 (33), 11210–11221. doi:
22
23 10.1523/JNEUROSCI.2239-10.2010.
24
25
26 De Lucia, M., Michel, C.M. and Murray, M.M., (2010b). Comparing ICA-based and single-trial
27
28 topographic ERP analyses. *Brain Topogr* 23 (2), 119–127. doi: 10.1007/s10548-010-
29
30 0145-y
31
32
33 De Meo, R., Murray, M.M., Clarke, S. and Matusz, P.J. (2015). Top-down control and early
34
35 multisensory processes: Chicken vs. egg. *Front Integr Neurosci*, 9, 17.
36
37
38 Dhamala, M., Assisi, C.G., Jirsa, V.K., Steinberg, F.L. and Kelso, J.A.S. (2007). Multisensory
39
40 intergation for timing engages different brain networks. *NeuroImage*, 34, 764–773.
41
42 doi: 10-1016/j.neuroimage.2006.07.044
43
44
45 Doehrmann, O., Weigelt, S., Altmann, C.F., Kaiser, J., and Naumer, M.J. (2010). Audiovisual
46
47 functional magnetic resonance imaging adaptation reveals multisensory integration
48
49 effects in object-related sensory cortices. *J Neurosci* 30(9), 3370-3379. doi:
50
51 10.1523/JNEUROSCI.5074-09.2010.
52
53
54
55 Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds
56
57 due to lip-reading. *Nature* 381 (6577), 66-68. doi: 10.1121/1.1907309.
58
59
60

- 1
2
3 Fairhall, S.L., and Macaluso, E. (2009). Spatial attention can modulate audiovisual integration
4
5 at multiple cortical and subcortical sites. *Eur J Neurosci*, 29, 1247-1257. doi:10.1111/
6
7 j.1460-9568.2009.06688.x
8
9
10 Fort, A., Delpuech, C., Pernier, J. and Giard M.H. (2002). Dynamics of cortico-subcortical
11
12 cross-modal operations involved in audio-visual object detection in humans. *Cereb*
13
14 *Cortex*, 12 (10), 1031–1039. doi: 10.1093/cercor/12.10.1031
15
16
17 Frost, R., Armstrong, B. C., Siegelman, N., and Christiansen, M. H. (2015). Domain generality
18
19 versus modality specificity: the paradox of statistical learning. *Trends Cogn Sci*, 19(3),
20
21 117-125. doi:10.1016/j.tics.2014.12.010
22
23
24 Froyen, D., van Atteveldt, N., Bonte, M., Blomert, L. (2009). Crossmodal enhancement of the
25
26 MMN to speech-sounds indicates early and automatic integration of letters and
27
28 speech-sounds. *Neurosci Lett*, 430, 23-28.
29
30
31 Goujon, A., and Fagot, J. (2013). Learning of spatial statistics in nonhuman primates:
32
33 contextual cueing in baboons (*Papio papio*). *Beh Brain Res*, 247, 101-109.
34
35 doi:10.1016/j.bbr.2013.03.004
36
37
38 Grahn, J.A., and Rowe, J.B. (2009). Feeling the beat: Premotor and striatal interactions in
39
40 musicians and non-musicians during beat perception. *J Neurosci*, 29(23), 7540-7548.
41
42 doi: 10.1523/JNEUROSCI.2018-08.2009
43
44
45 Grave de Peralta Menendez, R., Gonzalez Andino, S., Lantz, G., Michel, C.M., Landis, T.
46
47 (2001). Noninvasive localization of electromagnetic epileptic activity. I. Method
48
49 descriptions and simulations. *Brain Topogr*. 14 (2), 131–137.
50
51 doi: 10.1023/A:1012944913650
52
53
54
55 Grave de Peralta, R. and Gonzalez Andino S. (2002). Comparison of algorithms for the
56
57 localization of focal sources: Evaluation with simulated data and analysis of
58
59
60

- 1
2
3 experimental data. *Int J Bioelectromagn*, 4 (1).
4
5 Grave de Peralta Menendez, R., Murray, M.M., Michel, C.M., Martuzzi, R., Gonzalez Andino,
6
7 S.L., 2004. Electrical neuroimaging based on biophysical constraints. *Neuroimage*, 21
8
9 (2), 527–539. doi: 10.1016/j.neuroimage.2003.09.051
10
11
12 Guthrie, D. and Buchwald, J.S. (1991). Significance testing of difference potentials.
13
14 *Psychophysiology* 28 (2), 240–244. doi: 10.1111/j.1469-8986.1991.tb00417.x
15
16
17 Heinemann, A., Kunde, W., and Kiesel, A. (2009). Context-specific prime-congruency effects:
18
19 On the role of conscious stimulus representations for cognitive control. *Conscious*
20
21 *Cogn* 18, 966–976. doi: 10.1016/j.concog.2009.08.009
22
23
24 Henson, R., Shallice, T., and Dolan, R. (2000). Neuroimaging evidence for dissociable form of
25
26 repetition priming. *Science*, 287, 1269-1272. doi: 10.1126/science.287.5456.1269
27
28
29 Jiménez, L., and Méndez, C. (1999). Which attention is needed for implicit sequence
30
31 learning? *J Exp Psychol Learn Mem*, 25, 236–259. doi:10.1037/0278- 7393.25.1.236
32
33
34 King, J.A., Korb, F.M., Egner, T. (2012). Priming of control: implicit contextual cuing of top-
35
36 down attentional set. *J Neurosci* 32(24), 8192-8200. doi: 10.1523/JNEUROSCI.0934-
37
38 12.2012
39
40
41 Koenig, T. and Melie-Garcia, L. (2010). A method to determine the presence of averaged
42
43 event-related fields using randomization tests. *Brain Topogr* 23 (3), 233–242. doi:
44
45 10.1007/s10548-010-0142-1
46
47
48 Koenig, T., Stein, M., Grieder, M. and Kottlow, M. (2014). A tutorial on data driven methods
49
50 for statistically assessing ERP topographies. *Brain Topogr*, 27, 72-83. doi:
51
52 10.1007/s10548-013-0310-1
53
54
55 Kok, P., Rahnev, D., Jehee, J.F.M., Lau, H.C and de Lange, F. P (2012). Attention reverses the
56
57 effect of prediction in silencing sensory signals. *Cereb Cortex*, 22, 2197-2206. doi:
58
59
60

- 1
2
3 10.1093/cercor/bhr310
4
5 Larsson, J., & Smith, A.T. (2012). fMRI repetition suppression: neuronal adaptation or
6
7 stimulus expectation? *Cereb Cortex*, 22, 567-576. doi: 10.1093/cercor/bhr119
8
9
10 Lehmann, D. and Skrandies, W. (1980). Reference-free identification of components of
11
12 checkerboard evoked multichannel potential fields. *Electroen Clin Neuro*, 48, 609–
13
14 621. doi: 10.1016/0013-4694(80)90419-8
15
16
17 Macaluso, E., George, N., Dolan, R., Spence, C. and Driver, J. (2004). Spatial and temporal
18
19 factors during processing of audiovisual speech: a PET Study. *Neuroimage*, 21(2), 725-
20
21 732.
22
23
24 Marchant, J. L., Ruff, C. C., and Driver, J. (2012). Audiovisual synchrony enhances BOLD
25
26 responses in a brain network including multisensory STS while also enhancing
27
28 target-detection performance for both modalities. *Hum Brain Map*, 33(5), 1212-1224.
29
30 doi: 10.1002/hbm.21278
31
32
33 Matusz, P.J., Retsa, C., and Murray, M.M. (in reviews). Cross-modal activation of visual
34
35 cortex: Task-irrelevant yet not automatic.
36
37
38 Matusz, P.J., Broadbent, H., Ferrari, J., Forrest, B., Merkley, R., Scerif, G. (2015a). Multi-
39
40 modal distraction: Insights from children's limited attention. *Cognition*, 136, 156-165.
41
42 doi:10.1016/j.cognition.2014.11.031
43
44
45 Matusz, P.J., Thelen, A., Amrein, S., Geiser, E., Anken, J., Murray, M.M. (2015b). The role of
46
47 auditory cortices in the retrieval of single-trial auditory-visual object memories. *Eur J*
48
49 *Neurosci*, 41, 699-708. doi: 10.1111/ejn.12804.
50
51
52 Matusz, P.J. and Eimer, M. (2013). Top-down control of audiovisual search by bimodal search
53
54 templates. *Psychophysiology*, 50, 996-1009. doi: 10.1111/psyp.12086
55
56
57 Matusz, P.J. and Eimer, M. (2011). Multisensory enhancement of visual attentional capture
58
59
60

- 1
2
3 in visual search. *Psychon Bull Rev*, 18(5), 904-909. doi: 10.3758/s13423-011-0131-8.
4
5 Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L. and Grave de Peralta, R.
6
7 (2004). EEG source imaging. *Clin Neurophysiol*, 115, 2195–2222. doi:
8
9 10.1016/j.clinph.2004.06.001
10
11
12 Michel, C.M., Koenig, T., Brandeis, D., Gianotti, L.R. and Wackermann, J. (2009). *Electrical*
13
14 *neuroimaging*. Cambridge, UK: Cambridge UP.
15
16
17 Michel, C. M. and Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging
18
19 tool. *Neuroimage* 61, 371–385 doi 10.1016/j.neuroimage.2011.12.039
20
21
22 Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E and Foxe J.J. (2002)
23
24 Multisensory auditory-visual interactions during early sensory processing in humans:
25
26 a high-density electrical mapping study. *Cogn Brain Res* 14(1), 115-128. doi:
27
28 10.1016/S0926-6410(02)00066-6
29
30
31 Müller, N.G., Strumpf, H., Scholz, M., Baier, M., & Melloni, L. (2013). Repetition Suppression
32
33 versus Enhancement - It's Quantity That Matters. *Cereb Cortex*, 23, 315-322. doi:
34
35 10.1093/cercor/bhs009
36
37
38 Murray, M.M., Brunet, D. and Michel, C.M. (2008) Topographic ERP analyses: a step-by-step
39
40 tutorial review. *Brain Topogr* 20, 249 –264. doi 10.1007/s10548-008-0054-5
41
42
43 Murray, M. M., De Santis, L., Thut, G., and Wylie, G. R. (2009). The costs of crossing paths
44
45 and switching tasks between audition and vision. *Brain Cogn*, 69(1), 47-55.
46
47
48 Murray, M.M. and Wallace, M.T. (2012). *The Neural Basis of Multisensory Processes*. Boca
49
50 Raton, FL: CRC Press.
51
52
53 Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., & Matusz, P. J. (2015). The
54
55 multisensory function of the human primary visual cortex. *Neuropsychologia*.
56
57 doi:10.1016/j.neuropsychologia.2015.08.011
58 Perrin, F., Pernier, J., Bertrand, O. and
59
60

- 1
2
3 Echallier, J.F. (1989). Spherical splines for scalp potential and current density
4 mapping. *Electroen clin Neuro* 72(2), 184-187. doi:10.1016/0013-4694(89)90180-6
5
6
7
8 Powers, A.R., Hillock, A.R., and Wallace, M.T. (2009). Perceptual training narrows the
9 temporal window of multisensory binding. *J Neurosci*, 29(39), 12265–12274.
10 doi:10.1523/JNEUROSCI.3501-09.2009
11
12
13
14 Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clin*
15 *Neurophys*, 118(10), 2128-2148.
16
17
18
19
20 Pourtois, G., Delplanque, S., Michel, C. and Vuilleumier, P. (2008). Beyond conventional
21 event-related brain potential (ERP): Exploring the time-course of visual emotion
22 processing using topographic and principal component analyses. *Brain Topogr* 20 (4),
23 265–277. doi: 10.1007/s10548-008-0053-6
24
25
26
27
28
29 Recasens, M., Leung, S., Grimm, S., Nowak, R., & Escera, C. (2014). Repetition suppression
30 and repetition enhancement underlie auditory memory-trace formation in the
31 human brain: an MEG study. *Neuroimage*, doi: 10.1016/j.neuroimage.2014.12.031.
32
33
34
35
36 Rowland, B.A., and Stein, B.E. (2014). A model of the temporal dynamics of multisensory
37 enhancement. *Neurosci Biobeh Rev*, 41, 78–84. doi: 10.1016/j.neubiorev.2013.12.003
38
39
40
41 Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old
42 infants. *Science*, 274(5294), 1926-1928. doi: 10.1126/science.274.5294.1926
43
44
45
46 Sanabria, D., Soto-Faraco, S. and Spence, C. (2007). Spatial attention and audiovisual
47 interactions in apparent motion. *J Exp Psychol Hum Percept Perform*, 33 (4), 927-
48 937. doi:10.1037/0096-1523.33.4.927
49
50
51
52 Sanabria, D., Spence, C. and Soto-Faraco, S. (2007). Perceptual and decisional contributions
53 to audiovisual interactions in the perception of apparent motion: A signal detection
54 study. *Cognition*, 102, 299-310. doi: 10.1016/j.cognition.2006.01.003
55
56
57
58
59
60

- 1
2
3 Sandhu, R., and Dyson, B. J. (2013). Modality and task switching interactions using bi-modal
4
5 and bivalent stimuli. *Brain Cogn*, 82(1), 90-99. doi:10.1016/j.bandc.2013.02.011
6
7
8 Sarko, D. K., Nidiffer, A. R., Powers, I. I. I., A. R., Ghose, D., and Wallace, M. T. (2012). In: *The*
9
10 *Neural Basis of Multisensory Processes*, ed M. M. Murray and M. T. Wallace (Boca
11
12 Raton, FL: CRC Press), 191–215.
13
14
15 Sarmiento, B.R., Shore, D., Milliken, B., and Sanabria, D. (2012). Audiovisual interactions
16
17 depend on context of congruency. *Atten Percept Psychophys*, 74, 563-574. doi:
18
19 10.3758/s13414-011-0249-9
20
21
22 Segaert, K., Weber, K., de Lange, F.P., Petersson, K. M., and Hagoort P. (2013). The
23
24 suppression of repetition enhancement: A review of fMRI studies. *Neuropsychologia*,
25
26 51, 59–66 10.1016/j.neuropsychologia.2012.11.006
27
28
29 Spence, C., and Driver, J. (Eds.). (2004). *Crossmodal space and crossmodal attention*. Oxford,
30
31 UK: Oxford UP.
32
33
34 Stein, B.E. (2012). *The New Handbook of Multisensory Processing*. MIT Press: Cambridge,
35
36 MA.
37
38
39 Stevenson, R.A., Ghose, D., Fister, J.K., Sarko, D.K., Altieri, N.A., Nidiffer, A.R., Kurela, L.R.,
40
41 Siemann, J.K., James, T.W., and Wallace, M.T. (2014). Identifying and quantifying
42
43 multisensory integration: A tutorial review. *Brain Topogr*, 27(6), 707-730. doi:
44
45 10.1007/s10548-014-0365-7
46
47
48 Summerfield, C. and Egnér, T. (2009). *Expectation (and attention) in visual cognition*. *Trends*
49
50 *Cogn Sci*, 13(9), 403-409. doi: 10.1016/j.tics.2009.06.003
51
52
53 Talairach, J., and Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain: 3-
54
55 dimensional proportional system: an approach to cerebral imaging. G. Thieme
56
57 Medical Publishers, Stuttgart; New York.
58
59
60

- 1
2
3 Talsma, D. and Woldorff, M.G. (2005). Selective attention and multisensory
4
5 integration:multiple phases of effects on the evoked brain activity. *J Cogn Neurosci*
6
7 *17 (7)*, 1098-1114. doi:10.1162/0898929054475172
8
9
- 10 Talsma, D., Doty, T.J., Woldorff, M.G. (2007). Selective attention and audiovisual integration:
11
12 is attending to both modalities a prerequisite for early integration? *Cereb Cortex*, *17*,
13
14 679-690. doi:10.1093/ cercor/bhk016
15
16
- 17 Talsma, D., Senkowski, D., Soto-Faraco, S. and Woldorff M.G. (2010). The multifaceted
18
19 interplay between attention and multisensory integration. *Trends Cogn Sci* *14*, 400-
20
21 410. doi:10.1016/ j.tics.2010.06.008
22
23
- 24 Ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M.M., and Matusz, P.J.
25
26 (in review). Control mechanisms put COGs (Context-Object-Goals) into multisensory
27
28 processing.
29
30
- 31 Thelen, A., Cappe, C., and Murray, M. M. (2012). Electrical neuroimaging of memory
32
33 discrimination based on single-trial multisensory learning. *NeuroImage* *62*, 1478-
34
35 1488. doi: 10.1016/j.neuroimage.2012.05.027
36
37
- 38 Thelen, A., Matusz, P.J., and Murray, M.M. (2014). Multisensory context portends object
39
40 memory. *Curr Biol*, *15*, R734-735. doi: 10.1016/j.cub.2014.06.040.
41
42
- 43 Treisman, A. (1996). The binding problem. *Curr Opin Neurobiol*, *6*, 171-178. doi:
44
45 10.1016/S0959-4388(96)80070-5
46
47
- 48 Treisman, A. (2006). How the deployment of attention determines what we see. *Vis Cogn*, *14*
49
50 (4-8), 411-443. doi: 10.1080/13506280500195250
51
52
- 53 Tucker, D.M., Liotti, M., Potts, G.F., Russell, G.S., and Posner, M.I. (1994). Mapping the
54
55 working brain: Spatiotemporal analysis of brain electrical fields. *Hum Brain Mapp*, *1*,
56
57 134-152. doi: 10.1002/hbm.460010206
58
59
60

- 1
2
3 Turk-Browne, N.B., Yi, D.J., Leber, A.B., and Chun, M.M. (2007). Visual quality determines the
4
5 direction of neural repetition effects. *Cereb Cortex*, 17, 425-433. doi:
6
7 10.1093/cercor/bhj159
8
9
10 van Atteveldt, N., Murray, M.M., Thut, G., and Schroeder, C.E. (2014). Multisensory
11
12 integration: flexible use of general operations. *Neuron*, 81, 1240-1253. doi:
13
14 10.1016/j.neuron.2014.02.044
15
16
17 Von Kriegstein, K., and Giraud, A.L. (2006). Implicit multisensory associations influence voice
18
19 recognition. *PLoS Biol*, 4(10), e326.
20
21
22 Vossel, S., Mathys, C., Daunizeau, J., Bauer, M., Driver, J., Friston, K.J., Stephan, K.E. (2014).
23
24 Spatial attention, precision, and Bayesian inference: A study of saccadic response
25
26 speed. *Cereb Cortex*, 24(6), 1436-1450. doi: 10.1093/cercor/bhs418.
27
28
29 Wylie, G., and Allport, A. (2000). Task switching and the measurement of "switch costs". *Psyc*
30
31 *Res*, 63(3-4), 212-233. doi: 10.1007/s004269900003
32
33
34 Wylie, G.R., Murray, M.M., Javitt, D.C., and Foxe, J.J. (2009). Distinct neurophysiological
35
36 mechanisms mediate mixing costs and switch costs *J Cogn Neurosci*, 21(1), 105-118.
37
38
39 Yuval-Greenberg, S., and Deouell, L. Y. (2009). The dog's meow: asymmetrical interaction in
40
41 cross-modal object recognition. *Exp Brain Res*, 193(4), 603-14. doi: 10.1007/s00221-
42
43 008-1664-6.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

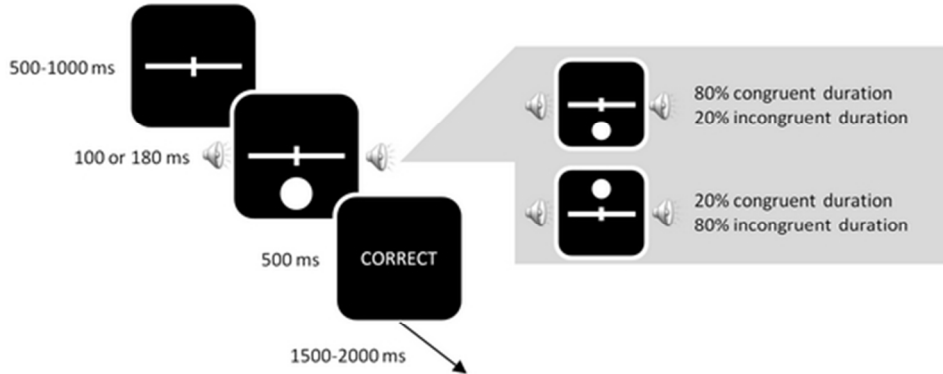
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Acknowledgements

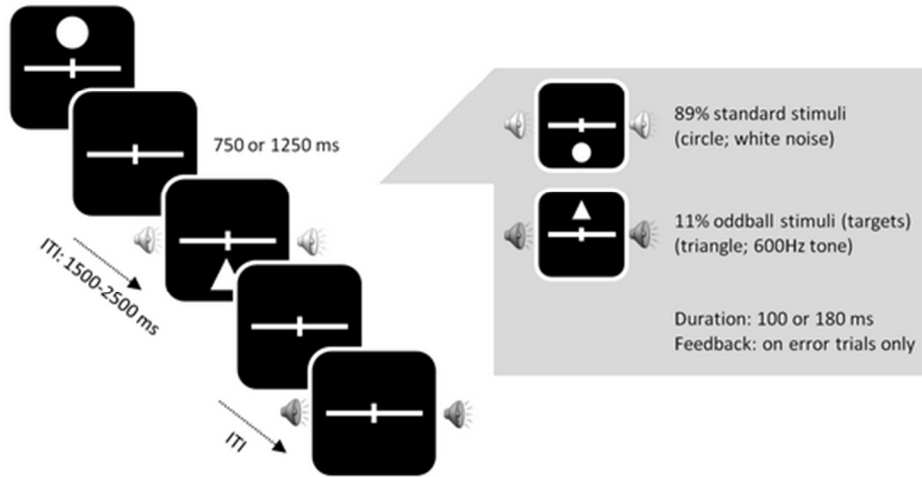
This study was supported by a Swiss National Science Foundation (grant 320030-149982 to M.M.M.), the Ministerio de Educación y Ciencia (student scholarship AP2008-03662 to B.R.S. and grant SEJ2007-63645 D.S.). We thank Diana Cutanda for her help with data collection and Antonia Thelen for her assistance with data analyses.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

a. Visual Temporal Discrimination Task

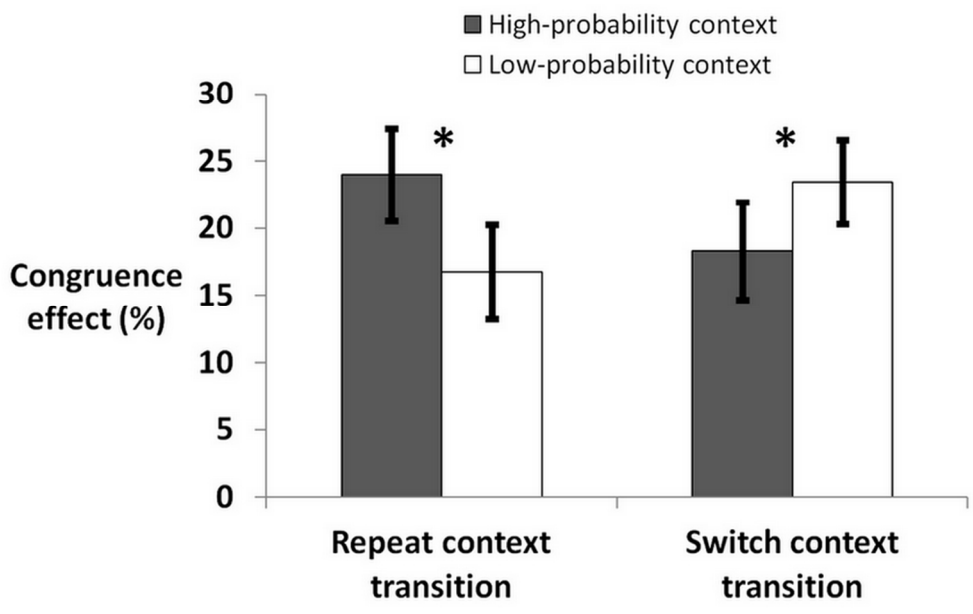


b. Oddball Task



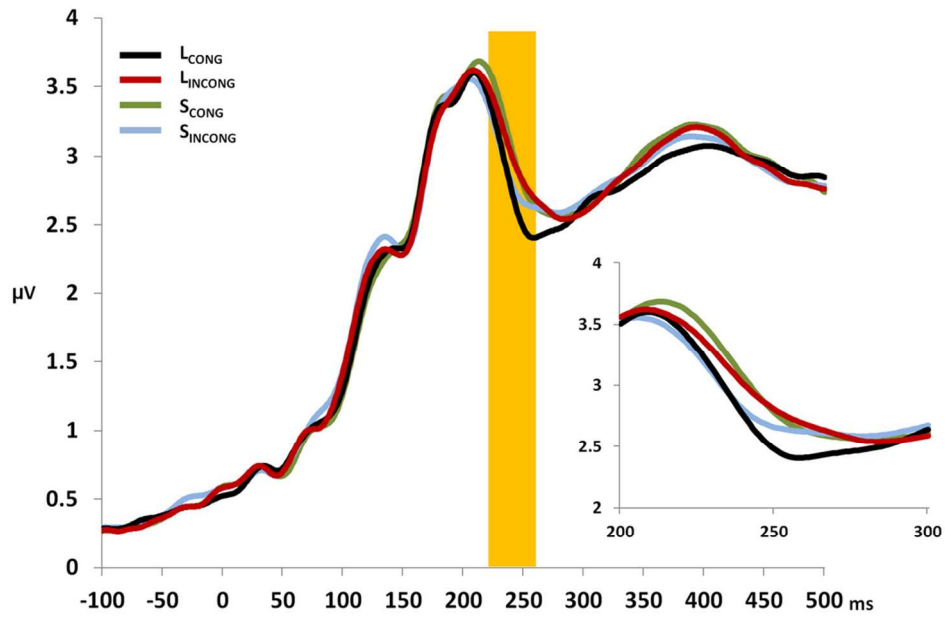
50x50mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



99x61mm (300 x 300 DPI)

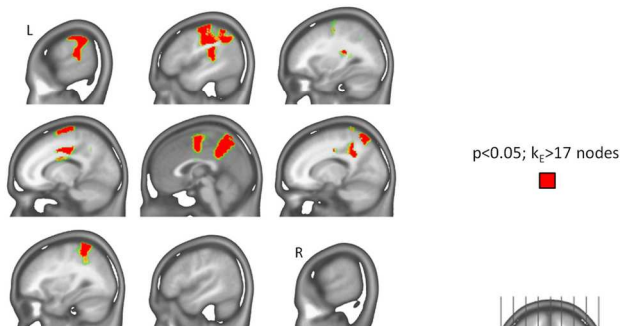
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



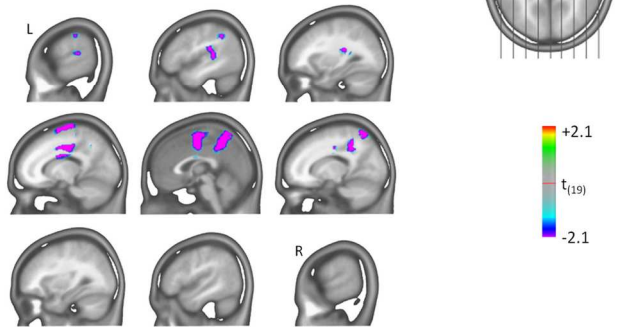
99x64mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

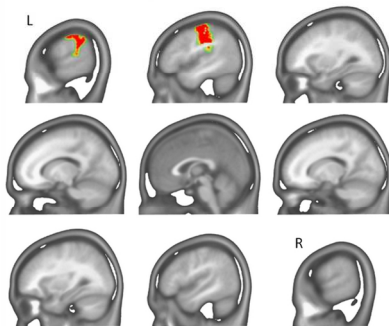
a. 2 Context of congruence x2 Congruence-effect size interaction (228-257 ms)



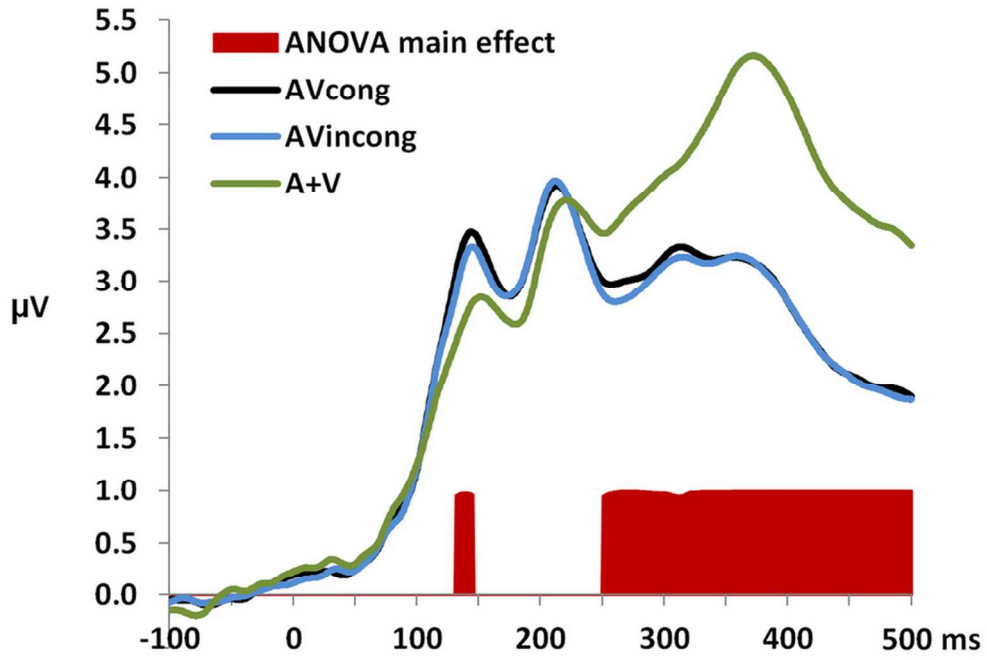
b. Large-size congruence-effect contrast



c. Small-size congruence-effect contrast



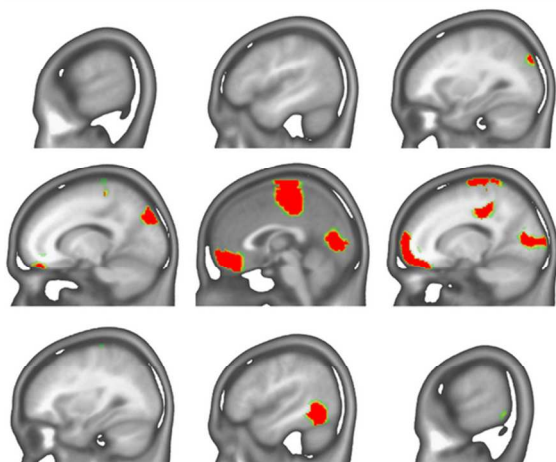
99x166mm (300 x 300 DPI)



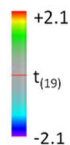
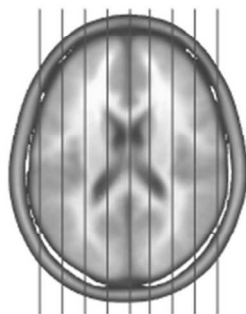
99x67mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

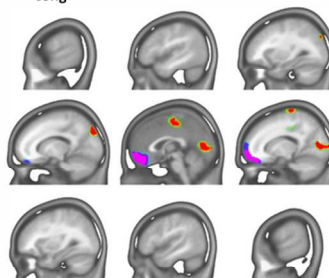
a. Main effect of stimulus condition



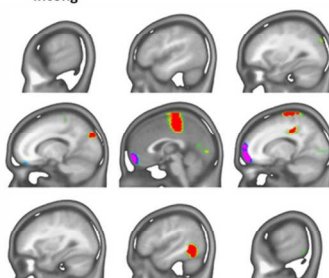
$p < 0.05$; $k_c > 17$ nodes



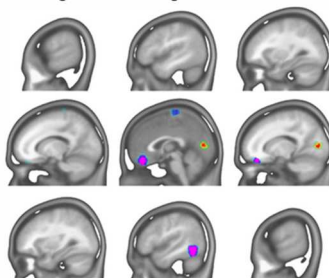
b. AV_{cong} vs. A+V



c. AV_{incong} vs. A+V



d. AV_{cong} vs. AV_{incong}



99x98mm (300 x 300 DPI)