

Serveur Académique Lausannois SERVAL [serval.unil.ch](http://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:** The role of auditory cortices in the retrieval of single-trial auditory-visual object memories.

**Authors:** Matusz PJ, Thelen A, Amrein S, Geiser E, Anken J, Murray MM

**Journal:** The European journal of neuroscience

**Year:** 2015 Mar

**Issue:** 41

**Volume:** 5

**Pages:** 699-708

**DOI:** [10.1111/ejn.12804](https://doi.org/10.1111/ejn.12804)

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.

## *Section: Cognitive Neuroscience*

# **The role of auditory cortices in the retrieval of single-trial auditory-visual object memories**

Running title: *Auditory cortices and single-trial object memories*

Pawel J. Matusz<sup>1,2,3,†,\*</sup>, Antonia Thelen<sup>1,4,†</sup>, Sarah Amrein, Eveline Geiser<sup>1</sup>, Jacques Anken<sup>1</sup>, Micah M. Murray<sup>1,4,5</sup>

<sup>1</sup>The Laboratory for Investigative Neurophysiology (The LINE), Department of Clinical Neurosciences and Department of Radiology, Vaudois University Hospital Center and University of Lausanne, Lausanne, Switzerland

<sup>2</sup>Attention, Behaviour, and Cognition Development Group, Department of Experimental Psychology, University of Oxford, United Kingdom

<sup>3</sup>University of Social Sciences and Humanities, Faculty in Wroclaw, Wroclaw, Poland

<sup>4</sup>Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>5</sup>Electroencephalography Brain Mapping Core, Center for Biomedical Imaging (CIBM) of Lausanne and Geneva, Switzerland

†Equal contributions

\*Address correspondence to:

Pawel J. Matusz  
Department of Clinical Neurosciences  
CHUV, BH07.071  
Rue du Bugnon 46  
1011 Lausanne, Switzerland  
[Pawel.matusz@chuv.ch](mailto:Pawel.matusz@chuv.ch)

Number of text pages	22
Number of figures	3
Number of appendices	0
Number of words in the whole manuscript	5,840
Number of words in the abstract	250
Number of words in the Introduction	499

*Keywords:* multisensory, cross-modal, event-related potential (ERP), human memory, auditory cortex

## **Abstract**

Single-trial encounters with multisensory stimuli affect both memory performance and early-latency brain responses to visual stimuli. Whether and how auditory cortices support memory processes based on single-trial multisensory learning is unknown and may moreover differ qualitatively and quantitatively from comparable processes within visual cortices due to purported differences in memory capacities across the senses. We recorded event-related potentials (ERPs) as healthy adults (N=18) performed a continuous recognition task in the auditory modality, discriminating initial (new) from repeated (old) sounds of environmental objects. Initial presentations were either unisensory or multisensory; the latter entailed synchronous presentation of a semantically congruent or a meaningless image. Repeated presentations were exclusively auditory, thus differing only according to the context in which the sound was initially encountered. Discrimination abilities (indexed by  $d'$ ) were increased for repeated sounds that were initially encountered with a semantically congruent image versus sounds initially encountered with either a meaningless or no image. Analyses of ERPs within an electrical neuroimaging framework revealed that early stages of auditory processing of repeated sounds were affected by prior single-trial multisensory contexts. These effects followed from significantly reduced activity within a distributed network, including the right superior temporal cortex, suggestive of an inverse relationship between brain activity and behavioural outcome on this task. The present findings demonstrate how auditory cortices contribute to long-term effects of multisensory experiences on auditory object discrimination. We propose a new framework for the efficacy of multisensory processes to impact both current multisensory stimulus processing and unisensory discrimination abilities later in time.

# 1. Introduction

Perception is inherently multisensory. Multisensory processes impact the initial stages of stimulus processing and continue onwards (e.g., Murray & Wallace, 2011; Stein, 2012; van Atteveldt et al., 2014). They affect, among other things, object detection and discrimination (Stein & Meredith, 1993; Murray et al., 2012) as well as attentional selection (Matusz & Eimer, 2011, 2013). Recent data demonstrated that multisensory processes occurring at one point in time also impact subsequent unisensory processes (Thelen & Murray, 2013; also Gibson & Maunsell, 1997; Nyberg et al., 2000; Wheeler et al., 2000; von Kriegstein & Giraud, 2006). While most studies have focused on the effects of multisensory context and learning on later visual processes, auditory processes are likewise affected. Discrimination of both images and sounds is improved for items originally encountered in a semantically congruent multisensory context, despite the multisensory information being task-irrelevant and only experienced in a single-trial setting (Thelen & Murray, 2013). Detriments or null effects on discrimination abilities can moreover be observed if the initial multisensory context had been either semantically incongruent or entailed a meaningless stimulus (Moran et al., 2013; Thelen & Murray, 2013; Thelen et al., 2014, in press). Overall, the behavioural effects of multisensory processing on auditory discrimination seem similar to those observed on visual discrimination.

Neuroimaging investigations elucidating brain mechanisms underlying multisensory processing affecting unisensory perception have largely focused on the effects of multisensory contexts on later unisensory visual processing. Object-sensitive visual cortices show differential responses during the first 100ms post-stimulus to visual stimuli that had been previously presented in a auditory-visual context (Murray et al., 2004; Thelen et al., 2012), with evidence for parallel effects within auditory cortices lacking (e.g., Murray et al., 2005). The effects of prior

multisensory experiences within visual cortices are not limited to tasks requiring visual discrimination. Voices and meaningless sounds both result in activations within visual cortices when previously studied in a multisensory context (fusiform face area: von Kriegstein & Giraud, 2006; medial occipital areas: Butler & James, 2011; Zangenehpour & Zatorre, 2010). It is unclear to what extent auditory cortices contribute to these effects.

We addressed the role of auditory cortices in the discrimination of unisensory auditory objects based on prior multisensory versus unisensory experiences. Resolving this issue will provide much-needed insights into the locus of the presumed multisensory object representation formed during the initial object encounter and the spatio-temporal neural substrates of its later access. One possibility is that auditory cortices do contribute to these processes, given the rapid cortical responses to sounds (Mussachia & Schroeder, 2009) and their semantic attributes (Bizley & Cohen, 2013; Murray & Spierer, 2009). Alternatively, multisensory representations may be predominantly located within visual cortices or rely on their co-activation with auditory regions, given the predominance of vision in object memories (Cohen et al. 2009; Yuval-Greenberg & Deouell, 2007, 2009). We differentiated between these possibilities via electrical neuroimaging of auditory ERPs recorded from adults performing a continuous old/new discrimination task involving sounds of environmental objects whose initial presentations entailed multisensory versus unisensory contexts.

## **2. Materials and Methods**

### ***2.1. Participants***

Twenty-two healthy adult volunteers (8 men) aged between 18 and 36 years old (mean  $\pm$  s.e.m. = 25  $\pm$  1 years) participated in the experiment. All subjects provided written, informed

consent for their participation. Experimental procedures were approved by the Ethics Committee of the Vaudois University Hospital Centre and the University of Lausanne and conformed to the World Medical Association Declaration of Helsinki (JAMA (2013), 310: 2191-2194). Data from four of these participants were excluded from analyses due to artefact contamination and/or a low number of accepted EEG epochs. The analyses presented here are based on the remaining eighteen participants (7 men; age range 22–36 years, mean age 25.9 years; 1 left-handed and 1 ambidextrous). All participants reported normal or corrected-to-normal vision and hearing and had no neurological or psychiatric illnesses. Data from a subset of these participants as well as subset of experimental conditions that are distinct from those analyzed here were presented as a follow-up experiment in Thelen et al. (2014).

## ***2.2. Stimuli and Procedure***

Participants performed a continuous recognition task involving the discrimination of initial versus repeated presentations of complex, meaningful sounds of environmental objects. This comprised a 2-alternative forced choice between old and new sounds. They were instructed to complete the task as fast and accurately as possible by a right-hand button press. When heard for the first time, sounds could be presented alone (A) or paired with either a semantically congruent image (AVc) or a meaningless image (AVm). Repeated sound presentations were always unisensory (auditory-only), and throughout the manuscript our nomenclature is according to their initial presentation context: previously presented alone (A-) or previously paired with a congruent (A+c) or meaningless image (A+m) (see Figure 1A). For the analyses, this led to six experimental conditions in total, though for task requirements there were only 2 conditions: initial and repeated presentations of sounds.

The sounds were of 60 common objects (obtained either from an online library [<http://dgl.microsoft.com>] or prior experiments (Marcell et al., 2000). These stimuli were modified with Adobe Audition to be of equal loudness (root mean square) and duration (500ms, including 10ms rise/fall to prevent clicks; 16 bit stereo, 44.1kHz digitization). Sounds were presented through stereo loud speakers (Logitech, Speaker System Z370) placed on each side of the monitor. The sound volume was adjusted to a comfortable level ( $53.3 \pm 0.2$ dB). We have previously used these sounds in studies of environmental object recognition and have controlled their familiarity to listeners (Murray et al., 2006). Nonetheless, to further ensure that the current participants understood to which object each sound referred, we presented each sound with its corresponding image twice during the course of a single block of trials preceding the main experiment. No responses were required of the participants.

For the AVc condition, visual stimuli were line drawings of common objects that were obtained either from a standardized set (Snodgrass & Vanderwart, 1980) or an online library ([dgl.microsoft.com](http://dgl.microsoft.com)). For the AVm condition, images were either abstract drawings (consisting of lines and circles, prepared by a professional artist) or scrambled versions of the line drawings (achieved using an in-house Matlab© script that divided the images into  $5 \times 5$  square matrices and randomized the arrangement of the pixels within each square). Visual stimuli were presented on a computer monitor (model HP LP2065; refresh rate of 60Hz) at a viewing distance of 120cm.

Stimuli were presented for 500ms followed by a variable inter-stimulus interval (ISI) of 1500 – 2500ms (mean=2000ms with uniform distribution). The experiment was divided into six blocks of 120 trials each. All six conditions were presented with the same likelihood over the 120 trials within an experimental block. Sounds were presented in initial encounter contexts in a counterbalanced way, such that over the six blocks each sound was presented twice in each initial

encounter context and was repeated once per block. There was an average of  $9\pm 4$  trials between initial and repeated presentations. The order of blocks was randomized between subjects. All stimuli were delivered and controlled by E-Prime 2.0, and responses were recorded with a serial response box (Psychology Software Tools, Pittsburgh, PA, USA).

### ***2.3. EEG acquisition and pre-processing***

Continuous EEG was recorded at 500Hz from 64 scalp electrodes (EasyCap, BrainProducts), positioned according to the international 10-20 system, using an electrode on the tip of the nose as the online reference. An additional EOG electrode was placed below the right eye. The mean impedances over all electrodes for all subjects were kept below  $4.5\text{k}\Omega$ . Continuous EEG was segmented into peri-stimulus epochs from 100ms pre-stimulus to 500ms post-stimulus onset. Only trials with correct responses were included in the analyses. Additionally, data quality was controlled with a semi-automated artefact rejection criterion of  $\pm 80\mu\text{V}$  at each channel as well as visual inspection to exclude transient noise and eye movements. To obtain event-related potentials (ERPs), the remaining epochs were averaged for all subjects for each of the three repetition conditions (A-, A+c, and A+m). The percentage of accepted epochs (out of a total of 120) per participant per condition was  $71.6\% \pm 2.9\%$  for A-,  $73.9\% \pm 3.2\%$  for A+c and  $72.0\% \pm 3\%$  for A+m. Data at artifact-contaminated or broken electrodes were interpolated using 3D splines (Perrin et al., 1987). On average, less than one electrode was interpolated per participant (range 0-3). ERPs were filtered with Butterworth 2<sup>nd</sup> order filters with -12dB/octave roll-off that were computed linearly with both forward and backward passes to eliminate any phase-shift (low-pass = 40Hz; high-pass = 0.1HzM; notch = 50Hz) and recalculated against the average reference

#### ***2.4. Behavioural Data Analyses***

Performance data were analysed using median reaction time (mRT) as well as according to signal detection theory, using perceptual sensitivity and response bias (Macmillan & Creelman, 2004). Perceptual sensitivity ( $d'$ ) was calculated using the hit rate (HIT = correct answer; e.g., labelling A+c as A+c) and the false alarm rate (FA = wrong discrimination; e.g., labelling AVc as A+c), according to the formula  $d' = [z(\text{HIT}) - z(\text{FA})]$ . Response bias ( $c$ ) was calculated according to the formula  $c = [0.5 * (z(\text{HIT})+z(\text{FA}))]$ . More generally, we would note the following regarding signal detection theory. Sensitivity ( $d'$ ) is a quantification of the distance between 2 (estimated) distributions: that of the signal and that of the noise. As such, it provides a measure of how easily a target signal is discriminated from the background noise. Response criterion (measured by  $c$ ) is a measure of how conservatively or liberally a participant makes a decision. These two measures have been widely characterized as reflecting perceptual vs. post-perceptual (i.e. decisional) stages (Macmillan & Creelman, 2004). Statistical analyses were conducted exclusively for the repeated sound presentations using a one-way analysis of variance (ANOVA) for the within-subject factor of condition (A-, A+c and A+m). Post-hoc t-tests were performed and p-values were corrected for multiple comparisons with the Holm-Bonferroni method (Holm, 1979). Throughout, we report mean  $\pm$  standard error of the mean. Analyses were performed using the SPSS software version 21 (International Business Machines Corporation, Armonk, NY, USA).

#### ***2.5. ERP Data Analyses***

ERP data were analysed within an electrical neuroimaging framework that focuses on reference-independent measures of the electric field at the scalp and uses multivariate analyses (Michel and Murray, 2012; Konig et al., 2014). We also performed univariate analyses of each scalp electrode as a function of time (1-way ANOVA for the within subject factor of condition). For this analysis we used an average reference as well as both a temporal criterion (>10ms contiguously; Guthrie and Buchwald, 1991) and spatial criterion (>5% of the electrode montage at a given latency) for the detection of statistically significant effects (see also Thelen et al., 2012, for a similar approach). This analysis is included to provide readers with a sense of the general waveform shape and the prototypical ERP components at the latency of our observed effects, although we emphasize that analyses of voltage waveforms are reference-dependent (Murray et al., 2008).

The electrical neuroimaging analyses entailed the following. As these methods have been described in several recent reviews (Koenig et al., 2014; Michel and Murray, 2012; Murray et al., 2008; Tzovara et al., 2012) and have been used extensively in ERP studies of sensation and perception (Altieri et al., 2013; Berchio et al., 2014; Chouiter et al., 2014; Hauthal et al., 2014; Hardmeier et al., 2014; Skrandies, 2014), we provide only the essential details here.

First, ERP strength was quantified using Global Field Power (GFP), which is reference-free and equals the root mean square across the electrode montage (Lehmann and Skrandies, 1980). GFP was analysed as a function of time using a 1-way ANOVA for the within-subject factor of condition (the same temporal criterion as above was applied).

Second, ERP topographic differences were quantified using Global Dissimilarity (DISS), which is reference-free and equals the root mean square of the difference between two GFP-normalized vectors (here the 64-channel ERP; Lehmann and Skrandies, 1980). DISS ranges in

value from 0 to 2, with 0 indicating no topographic differences and 2 indicating topographic inversion. DISS was analysed using the Randomization Graphical User interface (RAGU; Koenig et al., 2011). Briefly, RAGU performs a non-parametric randomization test on the DISS values, comparing the observed value to an empirical distribution based on permutations of the data from all participants/conditions. Topographic differences between conditions indicate that there is a difference in the configuration of the underlying neural generators (Lehmann, 1987). It should be noted that GFP and DISS are orthogonal measures of the ERP, which means that these two features (strength and topography) can be analysed independently.

Third, the collective post-stimulus group-average ERPs were subjected to a topographic analysis based on a hierarchical clustering algorithm (Murray et al. 2008). This clustering identifies stable electric field topographies (hereafter “template maps”). The clustering is exclusively sensitive to topographic modulations, because the data are first normalized by their instantaneous GFP. The optimal number of temporally-stable ERP clusters (i.e. the minimal number of maps that accounts for the greatest variance of the dataset) was determined using a modified Krzanowski-Lai criterion (Murray et al., 2008). The clustering makes no assumption regarding the orthogonality of the derived template maps (Pourtois et al. 2008; De Lucia et al. 2010a; Koenig et al., 2014). Template maps identified in the group-average ERPs were then submitted to a fitting procedure, wherein each time point of each single-subject ERP is labeled according to the template map with which it best correlated spatially (Murray et al. 2008). An ANOVA with within-subject factors of map and condition statistically tested the relative presence (in milliseconds) of each template map in the moment-by-moment scalp topography of the ERP and the differences in such across conditions.

Finally, we estimated the localization of the electrical activity in the brain using a distributed linear inverse solution (minimum norm) combined with the LAURA (local autoregressive average) regularization approach comprising biophysical laws as constraints (Grave de Peralta Menendez et al., 2001, 2004; see also Michel et al., 2004 for review). LAURA selects the source configuration that better mimics the biophysical behaviour of electric vector fields (i.e., activity at one point depends on the activity at neighbouring points according to electromagnetic laws). In our study, homogenous regression coefficients in all directions and within the whole solution space were used. LAURA uses a realistic head model, and the solution space included 4024 nodes, selected from a 6x6x6mm grid equally distributed within the gray matter of the Montreal Neurological Institute's average brain (courtesy of R. Grave de Peralta Menendez and S. Gonzalez Andino; <http://www.electrical-neuroimaging.ch/>). Prior basic and clinical research from members of our group and others has documented and discussed in detail the spatial accuracy of the inverse solution model used here (e.g., Gonzalez Andino et al. 2005; Grave de Peralta Menendez et al. 2004; Michel et al. 2004; Martuzzi et al., 2009). In general, the localization accuracy is considered to approximately follow the matrix grid size (here 6 mm). The results of the above topographic pattern analysis defined time periods for which intracranial sources were estimated and statistically compared across conditions. Prior to calculation of the inverse solution, the ERP data were down-sampled and affine-transformed to a 61-channel montage. Statistical analyses of source estimations were performed by first identifying the peak of the ERP over the time period demonstrating statistically reliably topographic differences (here, 32-84ms as detailed below). The inverse solution was then estimated for this peak latency for each of the 4024 nodes prior to conducting an ANOVA for the within-subjects factor of condition. Only solution points meeting the  $p \leq 0.05$  statistical criterion were considered

significant. Additionally, we applied a spatial extent criterion of at least 10 contiguous significant nodes (see also Toepel et al., 2009; Cappe et al., 2010, 2012; De Lucia et al., 2010b; Knebel et al., 2011, 2012 for a similar spatial criterion). This spatial criterion was determined using the AlphaSim program (available at <http://afni.nimh.nih.gov>) and assuming a spatial smoothing of 2mm full-width half maximum and cluster connection radius of 8.5mm. This criterion indicates that there is a 3.41% probability of a cluster of at least 10 contiguous nodes, which gives an equivalent node-level p-value of  $p \leq 0.00009$ . The results of the source estimations were rendered on the Montreal Neurologic Institute's average brain.

### **3. Results**

#### ***3.1. Behavioural data***

The overall mean accuracy on the task was  $74.3 \pm 2.3\%$ . To assess the impact of multisensory memories on the later recognition of environmental sounds, mRTs,  $d'$  and  $c$  were analysed for the repeated sound presentations. Participants performed the task equally fast in all conditions ( $F_{2,34} < 1$ ;  $965 \pm 27$ ms,  $965 \pm 29$ ms, and  $966 \pm 29$ ms for the A-, A+c, and A+m conditions, respectively). The ANOVA on the  $d'$  values revealed statistically reliable differences in participants' memory performance as a function of the initial context ( $F_{2,16} = 13.47$ ,  $P = 0.001$ ,  $\eta_p^2 = 0.63$ ). Post-hoc comparisons via paired t-tests revealed significantly better memory performance in the A+c condition ( $2.4 \pm 0.1$ ) than either the A- ( $2.0 \pm 0.1$ ,  $t_{17} = 3.69$ ,  $P = 0.0018$ , two-tailed) or the A+m condition ( $2.1 \pm 0.1$ ,  $t_{17} = 5.31$ ,  $P = 0.000058$ , two-tailed), the latter two of which did not significantly differ ( $t_{17} < 1$ , two-tailed; see Figure 1B). This pattern indicates that the discrimination of repeated sounds was higher in situations where past contexts were multisensory and semantically congruent. Importantly and by contrast,

participants showed no differences in response bias across the three conditions ( $F_{2,16} < 1$ ; Figure 1C), with values of  $0.31 \pm 0.09$ ,  $0.32 \pm 0.11$ , and  $0.31 \pm 0.12$  for the A-, A+c, and A+m conditions, respectively.

### 3.2. ERP data

As above, the ERP analyses focused on determining the effects of multisensory context on repeated sound discrimination. Figure 2 displays group-averaged ERPs from A-, A+c, and A+m conditions from an exemplar frontal midline electrode (Fz; Figure 2A) as well as the univariate ANOVA results across the electrode montage, displayed as the number of electrodes exhibiting a significant main effect of context as a function of time peri-stimulus and including a threshold of 10% of the electrode montage (Figure 2B). There were multiple time intervals exhibiting a main effect; the earliest of which was observed over the 30-52ms post-stimulus interval. Subsequent main effects were found over the 136-168ms, 224-274ms, 312-368ms, and 446-468ms time periods. While this analysis provides a general sense of the timing of ERP modulations, we focus the remainder of the Results section on the findings using an electrical neuroimaging framework (detailed in Materials and Methods).

The ANOVA on the GFP time series failed to document significant GFP differences at any latency during the -100 to 500ms post-stimulus period (data not shown). These results thus provided no evidence for differences in response strength as a function of past multisensory vs. unisensory experience. A set of topographic analyses were therefore conducted (see Methods for details). We would remind the reader that biophysical laws dictate that topographic differences forcibly follow from changes in the underlying configuration of active brain sources. First, the randomization-based analysis of the ERP topography (Koenig et al., 2011)

identified significant differences across the three conditions over the 32-50ms, 236-270ms, and 336-368ms post-stimulus periods. Second, topographic differences were likewise tested using a topographic cluster analysis of the full 500ms post-stimulus period. Eight clusters, involving 19 distinct template maps, accounted for 94.6% of the variance in the cumulative group-average ERPs. In general, the same pattern of template maps characterized the ERP in each condition, with the exception of the 32-84ms and 142-272ms post-stimulus periods (Figure 2C). We would remind the reader that a template map refers to a stable ERP topography observed in the group-averaged data that is then used for spatial correlation analyses at the single-subject level across all experimental conditions.

Over the 32-84ms post-stimulus period, the ERP in the A+c condition was dominated by one template map, whereas the ERPs in the A- and A+m conditions were both characterized by the same pair of template maps. This pattern observed in the group-averaged ERPs was statistically assessed in the single-subject data using a spatial-correlation fitting procedure. There was a significant condition x map interaction ( $F_{2,16} = 5.16$ ,  $P = 0.019$ ,  $\eta_p^2=0.39$ ), which was further confirmed by post-hoc comparisons likewise showing that one template map predominated ERPs elicited in the A+c condition versus either the A- or A+m condition (see Figure 2D).

Over the 142-272ms post-stimulus period, the ERP in the A+c condition was characterized by two template maps, whereas the ERPs in the A- and A+m conditions were both characterized by the same single template map. This pattern observed in the group-averaged ERPs was statistically assessed in the single-subject data using a spatial-correlation fitting procedure, as above. There was a non-significant trend for a condition x map interaction ( $F_{2,16} = 2.85$ ,  $P = 0.087$ ,  $\eta_p^2=0.26$ ). The robustness of this topographic difference will need confirmation

in future studies. Finally, over the 322-448ms post-stimulus period, the two template maps were observed across all conditions in the group-averaged ERPs, albeit with the appearance of differing relative durations. Statistical analysis based on spatial-correlation fitting did not yield a significant interaction between condition and map ( $F_{2,16} < 1$ ). Given the statistical reliability of the effects over the 32-84ms post-stimulus period across analyses of voltage waveforms and electric field topography as quantified both using global dissimilarity as well as clustering, we focused our source estimations on this early time period.

### ***3.3. Source Estimations***

Source estimations from the 32-84ms time period were statistically analysed to identify the likely brain regions contributing to these differential effects. The statistical contrast of these source estimations identified four clusters exhibiting a significant main effect of condition (Figure 3A). These clusters were located within the right posterior superior temporal cortex, the right inferior occipital cortex, the right inferior parietal cortex, and left frontal cortex. Overall, responses were significantly weaker for the A+c and A+m conditions than for the A- condition. In addition, responses were reliably weaker for the A+c than A+m condition within portions of the right posterior superior temporal cortex as well as right inferior parietal cortex. This pattern can be seen in Figure 3B displaying the mean scalar values at the node within each cluster exhibiting the largest F-value as well as in Figure 3C, which shows axial slices with the results of the post-hoc contrasts for pairs of conditions. Activity within the regions reported here essentially mirrors the pattern observed in behaviour as well as ERP analyses at the scalp surface; namely, responses to the A+c condition are distinct from those to either the A- or A+m conditions.

## 4. Discussion

The present study demonstrates that the neural recruitment of brain areas during early, sensory-perceptual stages of auditory object processing is strongly influenced by the sensory context in which a given auditory object was perceived initially. As predicted, discrimination of object sounds that had previously been encountered in a semantically congruent multisensory context was facilitated compared to sounds presented either alone or together with a meaningless image. Electrical neuroimaging analyses of ERPs elicited by the repeated sounds revealed that this facilitatory effect of past incidental encounters was accompanied by differential processing of the repeated sounds starting already at ~30ms post-stimulus onset, which followed from changes in the topography of the electric field map at the scalp. Source estimations localised these effects to right-hemisphere brain regions, most notably, the superior temporal cortex (STC), as well as the inferior parietal cortex (IPC). Overall, source activity was weaker for conditions where memory performance was improved. This is suggestive of an inverse relationship between the responsiveness of auditory cortices and the behavioural outcome in the ‘old/new’ task. The current findings are discussed in terms of the importance of auditory cortices in mediating supporting facilitated discrimination of auditory objects based on incidental single-trial multisensory learning.

### **The perceptual nature of single-trial multisensory learning**

Before treating in more depth our electrical neuroimaging results, the present behavioural findings warrant qualification. Previous work from our lab (Thelen et al., in press) provided some of the first evidence for the sensitivity of auditory object memory to single-trial multisensory memories: Sounds that had been initially presented together with a congruent image were

discriminated more accurately, as shown by higher sensitivity ( $d'$ ), but responded to equally quickly compared to sounds presented alone (or sounds presented with either a semantically incongruent or a meaningless, abstract image). More accurate discrimination of sounds that had been encountered in a multisensory (semantically congruent) vs. unisensory context was likewise recently demonstrated by Moran et al. (2013). However, despite a very similar paradigm, they found that sounds previously encountered in the multisensory context elicit slower responses. This contrasts with the typical finding of null effects on response times involving visual and auditory object recognition alike (for a review, see Thelen & Murray, 2013). The pattern observed by Moran et al. (2013) may thus follow from a speed-accuracy trade-off, raising doubts as to the mechanism that led to the observed increased performance accuracy. Here, we used analyses based on signal detection theory (Macmillan & Creelman, 2004) and demonstrated that facilitation of processing of semantically congruent multisensory pairings relative to sounds with unisensory past can take place without any effect on the response criterion. As in our prior studies, we failed to observe response time differences. Although we hesitate to over-interpret a null result, similarly to Thelen et al. (in press), we also found that sounds encountered initially without an image and those presented together with a meaningless image were discriminated with similar accuracy. Thus, the current behavioural results showing an effect on  $d'$  provide important support for the involvement of perceptual, rather than later, decisional processes, in single-trial multisensory learning and for the importance of this type of learning for auditory (and likely also visual) object processing. It is noteworthy that other forms of learning involving stimuli in multiple modalities (studied jointly or separately) typically require conscious effort and extensive periods of exposure/ training (in the range of hours; e.g., von Kriegstein & Giraud, 2006; Zangenehpour & Zatorre, 2010), which further underlines that the effects studied in our

continuous discrimination paradigm are engendered by a distinctive form of multisensory learning that is more implicit in nature.

### **The brain dynamics of semantic discrimination of sounds**

The timing of our electrical neuroimaging effects speaks to the perceptual nature of multisensory learning and memory at play in our task as well as, more generally, to the understudied role of auditory cortices in semantic processing. The fact that our effects were observed over the 30-85ms post-stimulus period could be misconstrued as these effects occurring too early for any semantic analysis. However, it is critical to situate their latency alongside current knowledge concerning signal propagation throughout the human auditory system, which is oftentimes extrapolated from the timing data from studies with non-human primates (using a 3:5 ratio; Musacchia & Schroeder, 2009). Response onset in the primary fields is on the order of 10-15ms (e.g., Steinschneider et al., 1982; reviewed in Musacchia & Schroeder, 2009). Intracranial recordings in humans indicate there is a synaptic delay of 2–3ms from primary to lateral superior temporal cortices (Brugge et al., 2003; Howard et al., 2000). Thus, there would be in principle another 15–75ms of cortical processing occurring prior to the effects at the latency observed in our study. This time would allow ample opportunity for recursive processing between superior temporal cortices and more or less remote loci, including frontal as well as medial temporal cortices, which are 1–2 synapses away (e.g., Kaas & Hackett, 2000; Romanski et al., 1999). What is more, complex sound discrimination that is semantic in nature manifests itself at latencies of ~30–100ms, as recorded intracranially in non-human primates within the superior temporal cortices (Russ et al., 2008; also Schnupp et al., 2006 for data from ferrets), the temporal pole (Ng et al., 2014), as well as the prefrontal cortices (e.g. Romanski et al., 2002). In humans,

EEG studies from our laboratory have likewise demonstrated semantic discrimination within the initial 70–100ms post-stimulus onset that was located within the superior and middle temporal cortices (e.g., De Lucia et al., 2010b; Murray et al., 2006). The timing and localisation of the present results are thus in strong agreement with the existing literature on the effects of semantic processing of sounds, while providing novel evidence that the semantic analysis of incoming auditory information can be influenced by its prior incidental and single-trial presentation in multisensory contexts.

### **Response suppression as a marker of efficient object discrimination**

Differential responses in the present study were characterized by significantly weaker activity for sounds that had been initially encountered in a multisensory rather than unisensory context. Moreover and within some regions showing this general multisensory context effect, activity was further reduced in response to sounds that had been initially encountered in a context involving a semantically congruent rather than a meaningless, abstract image. This type of response suppression is reminiscent of repetition suppression phenomena (De Lucia et al., 2010c; Desimone, 1996; Gibson & Maunsell, 1997; Murray et al., 2008). It has been argued that response suppression may reflect more efficient processing of a stimulus, involving activation of fewer and/or more selective units or populations (Grill-Spector et al., 2006). In agreement with this account, in a study using the classic delayed match-to-sample task Ng et al. (2014) provided evidence suggesting that response suppression may in fact be a characteristic signature of recognition memory within the auditory system. That the currently observed suppression effects transpired during early post-stimulus stages (i.e., within 100ms post-stimulus onset), and the strongest response suppression was observed for sounds that were discriminated most

successfully, is consistent with the notion that suppressed neural responses may reflect particularly efficient discrimination (e.g., Buchsbaum & D'Esposito, 2009; Rong et al., 2010). While a causal relationship between the specific patterns of neural activity and memory performance remains to be firmly established, to the extent that response suppression is an established proxy for behavioural outcomes would provide strong evidence for the efficacy of multisensory processes in memory functions, in audition as well as more generally. This notion is further supported by recent work from our group showing that the extent of multisensory processing during the initial encounter context is predictive of later memory performance under unisensory conditions (Thelen et al., 2014). Moreover, the present ERP results demonstrate that the mechanism of response suppression extends to a discrimination task involving more implicit memory processing, and, more importantly, to situations where memory performance varies according to the nature of the context of the initial encounter (multisensory vs. unisensory; semantically congruent vs. arbitrary, 'episodic' multisensory context; see also Thelen et al., 2012).

### **Linking performance and brain activity**

Prior observations provided a conflicting picture of the link between the occurrence (and direction) of activity changes in sensory cortices and memory performance. A prevailing account explaining consequences of multisensory contexts on memory performance is based on redintegration (Hamilton, 1859; James, 1890), which refers to the idea that a component of a consolidated experience (e.g., the voice of a previously-encountered person) is a sufficient cue to elicit responses within brain areas ordinarily activated by other components of that experience (e.g., the face of the same person). Neuroscientific investigations of this account have verified the

importance of such reactivation of modality-specific brain regions and support a traditional view of how redintegration would manifest; activity increases in brain areas associated with the non-stimulated modality (Butler & James, 2011; Nyberg et al., 2000; von Kriegstein & Giraud, 2006; Wheeler et al., 2000; see also Fuster, 2010). Together with previous results from our laboratory (for a review, see Thelen & Murray, 2013) the present findings suggest a novel perspective on how redintegration processes may operate, highlighting the likely importance of both task demands (i.e., explicit vs. implicit processing) as well as behavioural outcome (i.e., benefits vs. impairments).

With regard to the role of auditory cortices in engendering performance benefits based on multisensory memories, it is noteworthy that both STC and IPC exhibited the weakest activity in response to sounds that had been initially encountered in a semantically congruent multisensory context versus either a purely auditory context or multisensory context involving a meaningless image. These loci have been previously implicated in the integration of object features into unified representations (Beauchamp et al., 2004; Tanabe et al., 2005; Werner & Noppeney, 2010; see also Thelen et al. 2014). A contrasting viewpoint suggests that regions such as the superior temporal sulcus (STS) are not themselves the locus of the unified multisensory object representations, but rather that these regions, if anything, serve as a conduit for integrative processes elsewhere (Hocking & Price, 2008; Taylor et al. , 2006). In turn, others would suggest that multisensory representations of objects are predominantly localised to visual cortices, reflecting the presumed dominance of vision in object processes (e.g., Diaconescu et al., 2013; Molholm et al., 2004). However, in contrast with the latter two stances, even during a visual task auditory cortices were shown to be involved in discriminating between those images that had been previously encountered with vs. without sounds (Thelen et al., 2012).

More generally, the present results provide an important interpretational framework for these activation results; which hitherto were somewhat decoupled from behaviour. Several groups have shown differential responses in the STS or nearby auditory cortices to unisensory stimuli following multisensory learning or exposure. For example, both Nyberg et al. (2000) and Wheeler et al. (2000) reported increased activity within auditory cortices in response to words/labels studied together with versus without sounds. In the study by Wheeler et al. (2000) performance for all conditions was at ceiling levels, whereas in Nyberg et al. (2000), performance was *de facto* impaired by studying words with sounds, making it challenging to draw inferences on any direct links between brain responses and efficacy of multisensory learning. Further adding to the confusion regarding the role of auditory cortices in multisensory learning and memory, von Kriegstein and Giraud (2006) failed altogether to observe any reliable differences within temporal cortices as a function of whether the studied voices had been learned with a face versus name (though such effects were observed in the fusiform gyrus). The discrepant pattern of brain activity and performance outcomes observed across these studies is only exacerbated by the diversity of paradigms they utilised. Moreover, results from these studies could in principle be driven by mental imagery, particularly in those studies that employed tasks requiring explicit recall. By contrast, studies from our lab consistently used one paradigm with systematic manipulations (i.e., semantic congruence and general meaningfulness of the multisensory contexts; the task-relevant modality) to study single-trial and implicit processes and their effects on multisensory learning and memory performance. Our current and prior (Thelen et al., 2012) findings collectively show that there is an inverse relationship between the direction of responses within auditory cortices and memory performance. Specifically, the stronger the response of auditory cortices the worse is the observed memory performance. This suggestion is in strong

correspondence with the notion of effective stimulus processing via suppressed activity discussed earlier.

### **The role of non-auditory cortices**

It is noteworthy that response modulations within occipital and frontal cortices appeared to reflect sensitivity to the general multisensory vs. unisensory nature of the context of the initial exposure, rather than the nature of the multisensory context itself. That is, responses were significantly weaker to both the A+c and A+m conditions vs. the A- condition, with no significant difference between the former. This pattern would suggest that occipital and frontal regions respond differently to sounds depending on their initial encounter context but independently of its semantic congruency, and despite the multisensory nature of this context being task-irrelevant and occurring only on a single trial. A similar set of regions was also identified using magnetoencephalography in response to successful discrimination of test stimuli during an auditory delayed match-to-sample task involving tones (Rong et al., 2011).

While we can only speculate as to the root process(es) producing the effects in occipital and frontal cortices, one possibility would be that participants adopted a strategy of tagging incoming stimuli in a general manner as having been previously encountered in a multisensory context or not, irrespective of the nature of this context. However, this distinction would appear to occur in a fashion that does not directly affect memory performance; participants were able to benefit only from those multisensory contexts that were semantically congruent. To the extent this is the case, it is likewise noteworthy that response suppression was observed simultaneously across a wide network of brain regions (Figure 3), but profiles in only a subset of these mirrored the behavioural effects. This concomitant pattern would suggest there are parallel and

anatomically separable operations, some of which distinguish repeated stimuli according to their initial multisensory versus unisensory contexts, while others distinguish between semantically congruent and non-congruent multisensory contexts.

## **Conclusion**

Single-trial and task-irrelevant visual information was sufficient to dramatically impact early stages of subsequent auditory object processing, thus providing compelling evidence for the effectiveness of multisensory learning contexts. Our findings are a direct demonstration of the crucial contribution of auditory cortices to these effects, which unfold during early stages of stimulus processing, exhibit an inverse relationship with behavioural outcome, and transpire in a seemingly task-dependant manner (cf., Werner & Noppeney, 2010).

## **Acknowledgements**

Financial support was provided by the Swiss National Science Foundation (grants 320030 149982 to MMM and P2LAP3-151771 to AT) and by the Swiss Brain League (2014 Research Prize to MMM). The Cartool software has been programmed by Denis Brunet, from the Functional Brain Mapping Laboratory, Geneva, Switzerland, and is supported by the EEG Brain Mapping Core of the Center for Biomedical Imaging ([www.cibm.ch](http://www.cibm.ch)) of Geneva and Lausanne. It is freely available at: <http://sites.google.com/site/fbmlab/cartool/cartooldownload>. The authors declare no conflicts of interest.

## **Abbreviations**

ANOVA, Analysis of variance; DISS, Global Dissimilarity; ERPs, event-related potentials ; GFP, Global Field Power; IPC, inferior parietal cortex; LAURA, Local autoregressive average; mRT, median reaction time; RAGU, Randomization Graphical User interface; STC, superior temporal cortex; STS, superior temporal sulcus

## References

- Altieri, N., Stevenson, R.A., Wallace, M.T. & Wenger, M.J. (2015). Learning to associate auditory and visual stimuli: Behavioral and neural mechanisms. *Brain Topogr.*, **28**.
- Beauchamp, M. S., Lee, K.E., Argall, B.D. & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, **41**, 809–823.
- Berchio, C., Rihs, T.A., Michel, C.M., Brunet, D., Apicella, F., Muratori, F., Gallese, V. & Umiltà, M. A. (2014). Parieto-frontal circuits during observation of hidden and visible motor acts in children. A high-density EEG source imaging study. *Brain Topogr.*, **27**, 258–270.
- Bizley, J. K. & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.*, **14**, 693–707.
- Brugge, J.F., Volkov, I.O., Garell, P.C., Reale, R.A., & Howard III, M.A. (2003). Functional connections between auditory cortex on Heschl's gyrus and on the lateral superior temporal gyrus in humans. *J. Neurophys.*, **90**, 3750–3763.
- Buchsbaum, B.R. & D'Esposito, M. (2009). Repetition suppression and reactivation in auditory–verbal short-term recognition memory. *Cer. Cortex*, **19**, 1474–1485.
- Butler, A.J. & James, K.H. (2011). Cross-modal versus within-modal recall: Differences in behavioral and brain responses. *Beh. Brain Res.*, **224**, 387–396.
- Cappe, C., Thut, G., Romei, V. & Murray, M.M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J. Neurosci.*, **30**, 12572–12580.
- Cappe, C., Thelen, A., Romei, V., Thut, G. & Murray, M.M. (2012). Looming signals reveal Synergistic principles of multisensory integration. *J. Neurosci.*, **32**, 1171–1182.
- Chouiter, L., Dieguez, S., Annoni, J. M. & Spierer, L. (2014). High and low stimulus-driven conflict engage segregated brain networks, not quantitatively different resources. *Brain Topogr.*, **27**, 279–292.

- Cohen, M.A., Horowitz, T.S. & Wolfe, J.M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 6008–6010.
- De Lucia, M., Michel, C.M. & Murray, M.M. (2010a). Comparing ICA-based and single-trial topographic ERP analyses. *Brain Topogr.*, **23**, 119–127.
- De Lucia, M., Clarke, S. & Murray, M.M. (2010b). A temporal hierarchy for conspecific vocalization discrimination in humans. *J. Neurosci.*, **30**, 11210–11221.
- De Lucia, M., Cocchi, L., Martuzzi, R., Meuli, R.A., Clarke, S. & Murray, M.M. (2010c). Perceptual and semantic contributions to repetition priming of environmental sounds. *Cereb. Cortex*, **20**, 1676–1684.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 13494–13499.
- Diaconescu A.O., Hasher, L. & McIntosh, A.R. (2013). Visual dominance and multisensory integration changes with age. *Neuroimage*, **65**, 152–166.
- Fuster, J.M. (2010). Cortex and memory. *J. Cogn. Neurosci.*, **21**, 451–459.
- Gibson, J.R. & Maunsell, J.H. (1997). Sensory modality specificity of neural activity related to memory in visual cortex. *J. Neurophysiol.*, **78**, 1263–1275.
- Gonzalez Andino, S. L., Murray, M.M., Foxe, J.J. & Grave de Peralta Menendez, R. (2005). How single-trial electrical neuroimaging contributes to multisensory research. *Exp. Brain Res.*, **166**, 298–304.
- Grave de Peralta Menendez, R., Gonzalez Andino, S., Lantz, G., Michel, C.M. & Landis, T. (2001). Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. *Brain Topogr.*, **14**, 131–137.
- Grave de Peralta Menendez, R., Murray, M.M., Michel, C.M., Martuzzi, R. & Gonzalez Andino, S. L. (2004). Electrical neuroimaging based on biophysical constraints. *Neuroimage*, **21**, 527–539.
- Grill-Spector, K., Henson, R. & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cog. Sci.*, **10**, 14–23.

- Guthrie, D. & Buchwald, J.S. (1991). Significance testing of difference potentials. *Psychophysiology*, **28**, 240–244.
- Hamilton, W. (1859). *Lectures on Metaphysics and Logic*. Gould & Lincoln, Boston (MA).
- Hauthal, N., Thorne, J.D., Debener, S. & Sandmann, P. (2014). Source localisation of visual evoked potentials in congenitally deaf individuals. *Brain Topogr.*, **27**, 412–424.
- Hardmeier, M., Hatz, F., Naegelin, Y., Hight, D., Schindler, C., Kappos, L., Seeck, M., Michel, C.M. & Fuhr, P. (2014). Improved characterization of visual evoked potentials in multiple sclerosis by topographic analysis. *Brain Topogr.*, **27**, 318–327.
- Hocking, J. & Price, C. J. (2008). The role of the posterior superior temporal sulcus in audiovisual processing. *Cer. Cortex*, **18**, 2439–2449.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, **6**, 65–70.
- Howard, M.A., Volkov, I.O., Mirsky, R., Garell, P.C., Noh, M.D., Granner, M., Damasio, H., Steinschneider, M., Reale, M.A., Hind, J.E. & Brugge, J.F. (2000). Auditory cortex on the human posterior superior temporal gyrus. *J. Comp. Neurol.*, **416**, 79–92.
- James, W. (1890). *Principles of Psychology*. Holt, New York (NY).
- Kaas, J.H. & Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 11793–11799.
- Knebel, J.F., Javitt, D.C. & Murray, M.M. (2011). Impaired early visual response modulations to spatial information in chronic schizophrenia. *Psychiatry Res.*, **193**, 168–176.
- Knebel, J.F. & Murray, M.M. (2012). Towards a resolution of conflicting models of illusory contour processing in humans. *Neuroimage*, **59**, 2808–2817.
- Koenig, T., Kottlow, M., Stein, M. & Melie-García, L. (2011). RAGU: A free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Computat. Intell. Neurosci.*, **2011**, 1–14.

- Koenig, T., Stein, M., Grieder, M. & Kottlow, M. (2014). A tutorial on data-driven methods for statistically assessing ERP topographies. *Brain Topogr.*, **27**, 72–83.
- Lehmann, D. & Skrandies, W., (1980). Reference-free identification of components of checkerboard evoked multichannel potential fields. *Electroencephalogr. Clin. Neurophysiol.*, **48**, 609– 621.
- Lehmann, D. (1987). Principles of spatial analysis. In Gevins, A.S., Remond, A. (Eds.), *Handbook of Electroencephalography and Clinical Neurophysiology. Methods of Analysis of Brain Electrical and Magnetic Signals, vol. 1*. Elsevier, Amsterdam, pp. 309–354.
- Macmillan, N.A. & Creelman, C.D. (2004). *Detection Theory: A User's Guide*. Psychology Press, Mahwah, New Jersey (NJ).
- Marcell, M.M., Borella, D., Greene, M., Kerr, E. & Rogers, S. (2000). Confrontation naming of environmental sounds. *J. Clin. Exp. Neuropsychol.*, **22**, 830–864.
- Martuzzi, R., Murray, M.M., Meuli, R.A., Thiran, J.P., Maeder, P.P., Michel, C.M., Grave de Peralta Menendez, R. & Gonzalez Andino, S.L. (2009). Methods for determining frequency- and region-dependent relationships between estimated LFPs and BOLD responses in humans. *J. Neurophysiol.*, **101**, 491–502.
- Matusz, P.J. & Eimer, M. (2011). Multisensory enhancement of visual attentional capture in visual search. *Psychon. B. Rev.*, **18**, 904–909.
- Matusz, P.J. & Eimer, M. (2013). Top-down control of audiovisual search by bimodal search templates. *Psychophysiology*, **50**, 996–1009.
- Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L. & Grave de Peralta Menendez, R. (2004). EEG source imaging. *Clin. Neurophysiol.*, **115**, 2195–2222.
- Michel, C.M. & Murray, M.M. (2012). Towards the utilization of EEG as a brain imaging tool. *Neuroimage*, **61**, 371–385.
- Molholm, S., Ritter, W., Javitt, D.C. & Foxe, J.J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cer. Cortex*, **14**, 452–465.

- Moran, Z.D., Bachman, P., Pham, P., Cho, S.H., Cannon, T.D. & Shams, L. (2013). Multisensory encoding improves auditory recognition. *Multisensory Res.*, **26**, 581–592
- Murray, M.M., Michel, C.M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S. & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage*, **21**, 125–135.
- Murray, M.M., Foxe, J.J. & Wylie, G.R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. *Neuroimage*, **27**, 473–478.
- Murray, M.M., Camen, C., Gonzalez, Andino S.L., Bovet, P. & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *J Neurosci*, **26**, 1293–1302.
- Murray, M.M., Brunet, D. & Michel, C.M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr*, **20**, 249–264.
- Murray, M.M. & Spierer, L. (2009). Auditory spatio-temporal brain dynamics and their consequences for multisensory interactions in humans. *Hearing Res.*, **258**, 121–133.
- Murray, M. M. & Wallace, M.T. (Ed.) (2011). *The Neural Bases of Multisensory Processes*. Taylor & Francis Group, Boca Roton, London, New York.
- Murray, M.M., Cappe, C., Romei, V., Martuzzi, R. & Thut, G. (2012). Auditory-visual multisensory interactions in humans: a synthesis of findings from behavior, ERPs, fMRI, and TMS. In Stein, B.E. (Ed.) *The New Handbook of Multisensory Processes*. MIT Press, Cambridge (MA), pp. 223–238.
- Musacchia, G. & Schroeder, C.E. (2009). Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hearing Res.*, **258**, 72–79.
- Ng, C. W., Plakke, B. & Poremba, A. (2014). Neural correlates of auditory recognition memory in the primate dorsal temporal pole. *J. Neurophysio.*, **111**, 455-469.
- Nyberg, L., Habib, R., McIntosh, A.R. & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 11120-11124.

- Perrin, F., Pernier, J., Bertrand, O., Giard, M.H. & Echallier, J.F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalogr. Clin. Neurophysiol.*, **66**, 75-81.
- Pourtois, G., Delplanque, S., Michel, C. & Vuilleumier, P. (2008). Beyond conventional event-related brain potential (ERP): Exploring the time-course of visual emotion processing using topographic and principal component analyses. *Brain Topogr.*, **20**, 265–277.
- Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S. & Rauschecker, J.P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci.*, **2**, 1131–1136.
- Rong, F., Holroyd, T., Husain, F.T., Contreras-Vidal, J.L. & Horwitz, B. (2010). Task-specific modulation of human auditory evoked response in a delayed-match-to-sample task. *Front. Psych.*, **2**, 1–16.
- Russ, B.E., Ackelson, A.L., Baker, A.E. & Cohen, Y.E. (2008). Coding of auditory-stimulus identity in the auditory non-spatial processing stream. *J. Neurophysio.*, **99**, 87–95.
- Schnupp, J.W., Hall, T.M., Kokelaar, R.F. & Ahmed, B. (2006). Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J. Neurosci.*, **26**, 4785–4795.
- Skrandies, W. (2014). Electrophysiological correlates of connotative meaning in healthy children. *Brain Topogr.*, **27**, 271–278.
- Snodgrass, J. G. & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Mem. Learn.*, **6**, 174–215.
- Stein, B. E. (Ed.). (2012). *The New Handbook of Multisensory Processing*. MIT Press, Cambridge (MA).
- Steinschneider, M., Tenke, C.E., Schroeder, C.E., Javitt, D.C., Simpson, G.V., Arezzo, J.C. & Vaughan Jr., H.G. (1992). Cellular generators of the cortical auditory evoked potential initial component. *Electroen. Clin. Neuro.*, **84**, 196–200.

- Tanabe, H. C., Honda, M. & Sadato, N. (2005). Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. *J. Neurosci*, **25**, 6409–6418.
- Taylor, K.I., Moss, H.E., Stamatakis, E.A., & Tyler, L.K. (2006). Binding crossmodal object features in perirhinal cortex. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 8239–8244.
- Thelen, A., Cappe, C. & Murray, M.M. (2012). Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *Neuroimage*, **62**, 1478–1488.
- Thelen, A. & Murray, M.M. (2012). The efficacy of single-trial multisensory memories. *Multisensory Res.*, **26**, 483–502.
- Thelen, A., Matusz, P.J. & Murray, M.M. (2014). Multisensory context portends object memory. *Curr. Biol.*, **24**, R734–R735.
- Thelen, A., Talsma, D. & Murray, M.M. (in press). Single-trial multisensory memories affect later visual and auditory object discrimination. *Cognition*.
- Toepel, U., Knebel, J. F., Hudry, J., le Coutre, J. & Murray, M.M. (2009). The brain tracks the energetic value in food images. *Neuroimage*, **44**, 967–974.
- Tzovara A, Murray M.M., Michel C.M. & De Lucia M. (2012). A tutorial review of electrical neuroimaging from group-average to single-trial event-related potentials. *Dev. Neuropsychol.*, **37**, 518–544.
- van Atteveldt, N., Murray, M.M., Thut, G. & Schroeder, C.E. (2014). Multisensory integration: flexible use of general operations. *Neuron*, **81**, 1240–1253.
- von Kriegstein, K. & Giraud, A.L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biol.*, **4**, e326.
- Werner, S. & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J. Neurosci.*, **30**, 2662–2675.
- Wheeler, M.E., Petersen, S.E. & Buckner, R.L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 11125–11129.

Yuval-Greenberg, S. & Deouell, L. Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *J. Neurosci.*, **27**, 1090–1096.

Yuval-Greenberg, S. & Deouell, L.Y. (2009). The broadband-transient induced gamma-band response in scalp EEG reflects the execution of saccades. *Brain Topogr.*, **22**, 3–6.

Zangenehpour, S. & Zatorre, R.J. (2010). Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia*, **48**, 591–600.

## Figure legends

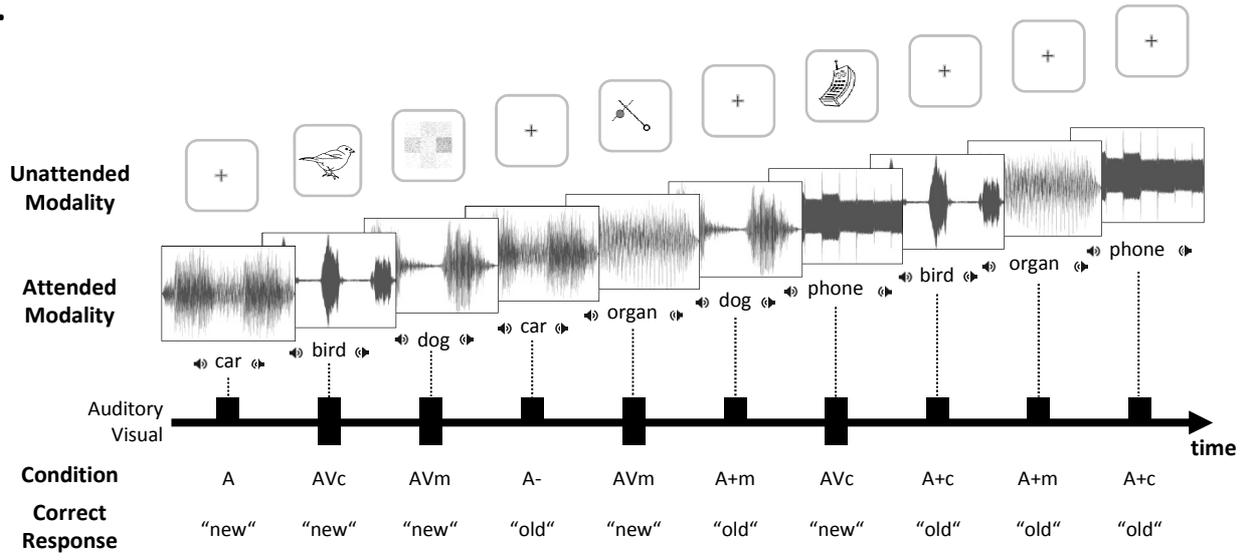
**Figure 1. A.** Illustration of the paradigm used to investigate the effects of multisensory learning on auditory object discrimination. **B and C.** Group-averaged sensitivity,  $d'$ , and response criterion,  $c$ , for the three experimental conditions: repeated presentations of auditory stimuli initially encountered in a purely auditory context (A-), semantically congruent multisensory context (A+c), and a meaningless multisensory context (A+m). Error bars indicate standard error of the mean. Significant post-hoc effects ( $P < 0.05$ ) are marked with an asterisk.

**Figure 2. A.** Group-averaged (s.e.m. indicated) auditory evoked potentials at an exemplar frontal midline electrode shown separately for the three repeated sound presentation conditions. **B.** The results of the millisecond-by-millisecond 1-way ANOVA across the electrode montage displaying the number of electrodes showing a main effect of condition and meeting the  $P < 0.05$  criterion for over 10ms contiguously and across at least 10% of the electrode montage (shaded region). **C.** The results of the millisecond-by-millisecond 1-way ANOVA on the strength-normalized electric field topography showing a main effect of condition and meeting the  $P < 0.05$  criterion for over 10ms contiguously. **D.** The results of the topographic cluster analysis identified two template maps over the 34-86 ms post-stimulus interval that were then used for the single-subject fitting procedure. Note that both template maps are characterized by a fronto-central positivity and that the intensity of these template maps is

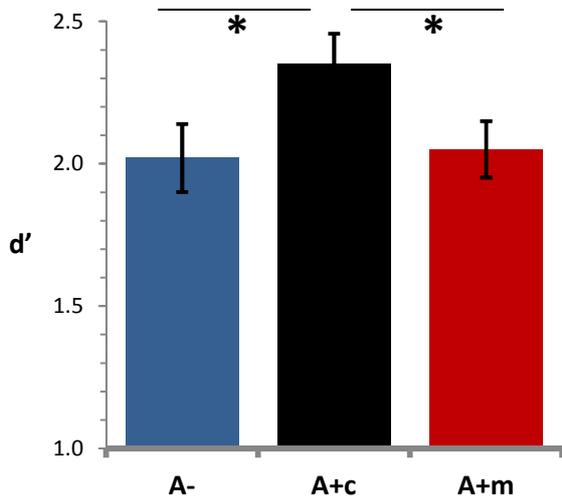
arbitrary given that this analysis entailed strength-normalized data to isolate topographic differences. The single-subject fitting results show the group-average duration each template map was ascribed to each condition (s.e.m. indicated). The 2x3 ANOVA on these duration values times revealed an interaction between template map and condition. The post-hoc comparisons showed this interaction was due to the predominance of one template map in the A+c condition. This was not observed for either the A- or A+m conditions.

**Figure 3. A.** The results of the 1-way ANOVA on source estimations calculated over the 34-86 ms post-stimulus interval showed a significant main effect of condition within a distributed set of brain regions, indicated by numbers. Data are shown on axial slices with the left hemisphere on the left and the nasion upwards. **B.** Within each region depicted in panel **A** are shown the mean current source density values at the locus of maximal F-value for each condition (s.e.m. indicated). **C.** Post-hoc comparisons show t-values specifically for nodes within the right posterior superior temporal cortex.

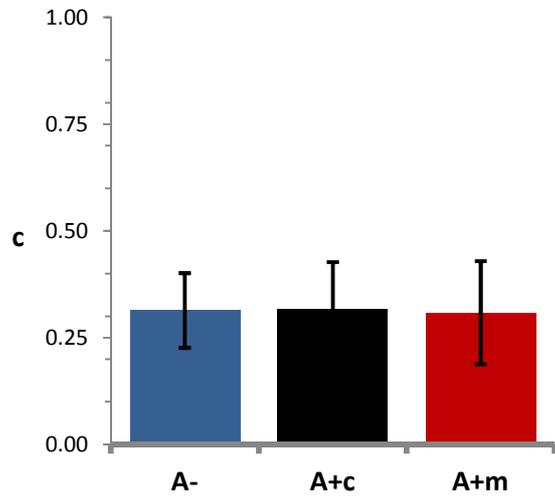
**A.**

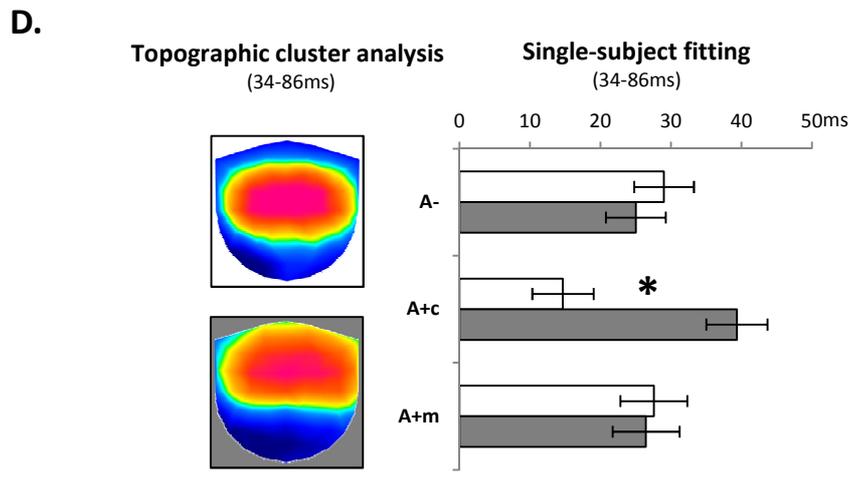
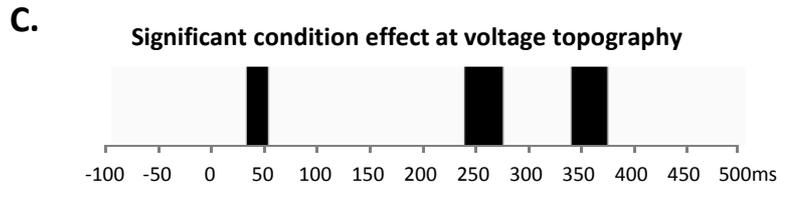
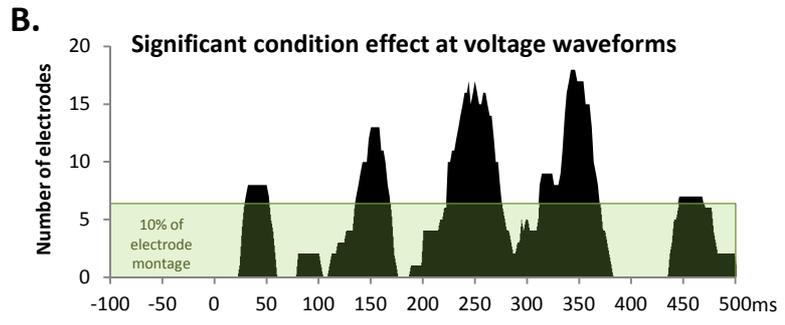
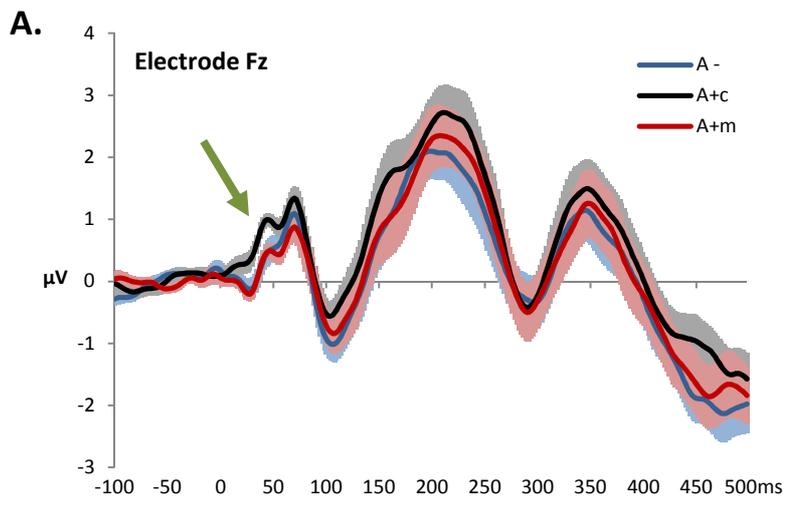


**B.**

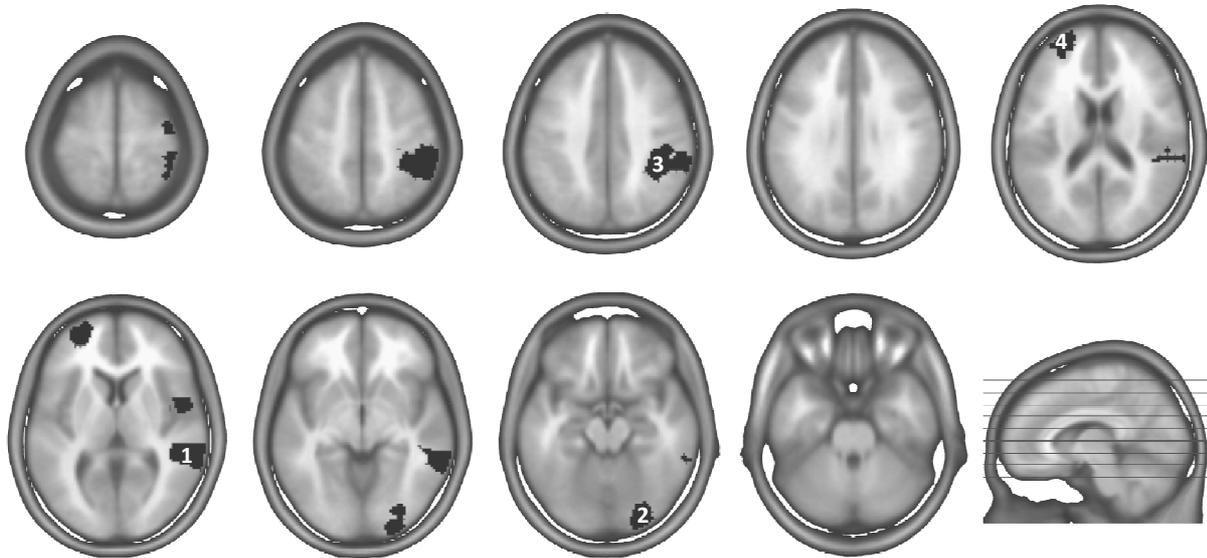


**C.**

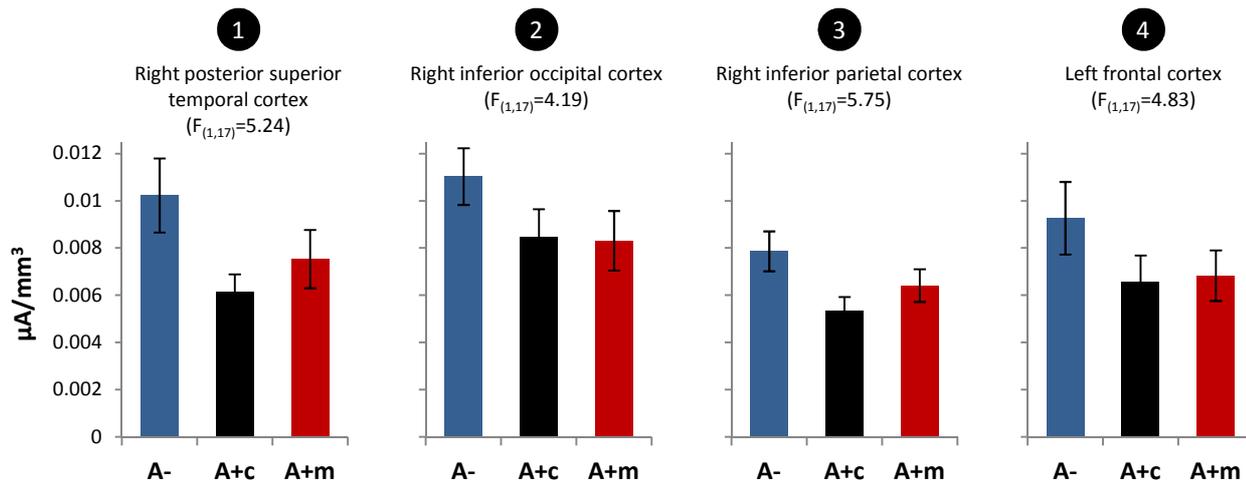




### A. Loci of significant condition effect



### B. Scalar value at locus of maximal F-value within each cluster



### C. Paired contrasts

