

Visual cortex signals a mismatch between regularity of auditory and visual streams

Running Title: *Mismatch responses to multisensory regularity*

Michael Andric¹, Ben Davis¹, & Uri Hasson^{1,2}

¹ Center for Mind/Brain Sciences (CIMEC), University of Trento, Italy.

² Center for Practical Wisdom, The University of Chicago

This is pre-print version of manuscript in press in *NeuroImage*

Address correspondence to:

Michael Andric
Center for Mind/Brain Sciences
Via delle Regole, 101
38123 Mattarello (TN)
Italy.
Email: andric.michael@gmail.com
Phone: +39-3203055679

Acknowledgment: This project was supported by European Research Council Starting Grant ERC-STG #263318 NeuroInt to U. H.,

Abstract

Understanding how humans code for and respond to environmental uncertainty/regularity is a question shared by current computational and neurobiological approaches to human cognition. To date, studies investigating neurobiological systems that track input uncertainty have examined responses to uni-sensory streams. It is not known, however, whether there exist brain systems that combine information about the regularity of input streams presented to different senses. We report an fMRI study that aimed to identify brain systems that relate statistical information across sensory modalities. We constructed temporally extended auditory and visual streams, each of which could be random or highly regular, and presented them concurrently. We found strong signatures of “regularity matching” in visual cortex bilaterally; responses were higher when the level of regularity in the auditory and visual streams mismatched than when it matched, [(AudHigh/VisLow and AudLow/VisHigh) > (AudLow/VisLow and AudHigh/VisHigh)]. In addition, several frontal and parietal regions tracked regularity of the auditory or visual stream independently of the other stream’s regularity. An individual-differences analysis showed that signatures of single-modality-focused regularity tracking in these fronto-parietal regions are inversely related to signatures of regularity-matching in visual cortex. Our findings suggest that i) visual cortex is a junction for integration of temporally-extended auditory and visual inputs and that ii) multisensory regularity-matching depends on balanced processing of both input modalities. We discuss the implications of these findings for neurobiological models of uncertainty and for understanding computations that underlie multisensory interactions in occipital cortex.

1. Introduction

Recent neurobiological research on statistical learning, decision-making, and coding of regularity has identified neural systems whose activity tracks input uncertainty. Studies relying on visual stimuli generally implicate the hippocampus and front-parietal systems in sensitivity to statistical structure (Harrison, Bestmann, Rosa, Penny, & Green, 2011; Huettel, Mack, & McCarthy, 2002; Strange, Duggins, Penny, Dolan, & Friston, 2005; Turk-Browne, Scholl, Johnson, & Chun, 2010) or prediction (Egner et al., 2008). In contrast, studies examining auditory stimuli have tended to implicate lateral temporal and inferior frontal regions (e.g., Cunillera et al., 2009; Karuza et al., 2013; McNealy, Mazziotta, & Dapretto, 2006; Nastase, Iacovella, & Hasson, 2014; Tobia, Iacovella, & Hasson, 2012; Tremblay, Baroni, & Hasson, 2013), with some implicating the basal ganglia (e.g., Geiser, Notter, & Gabrieli, 2012; McNealy, Mazziotta, & Dapretto, 2006). Some lateral-temporal regions also track sub-lexical statistics in natural language (e.g., Leonard, Bouchard, Tang, & Chang, 2015; Tremblay, Deschamps, Baroni, & Hasson, 2016). All this suggests that regularities of different sensory streams may be tracked in different neural systems (see Dehaene et al., 2015, for review). Such results are consistent with behavioral findings, which yield little evidence for a single latent factor underlying sensitivity to statistics in auditory and visual streams (Siegelman & Frost, 2015). It also suggests that a domain-general capacity, if existent, would be subservient to modality-specific processing constraints (Frost, Armstrong, Siegelman, & Christiansen, 2015). However, all aforementioned studies share a core feature: they develop from, and evaluate theoretical models based on input streams presented within just a single modality – typically either auditory or visual.

What is not known is how the brain responds to the more complex and ordinary case in which statistical information is concurrently available in multiple modalities. For instance, weather patterns are associated with particular arrangements of temporally unfolding visual and auditory phenomena, as are warehouses, airplanes, and urban environments. A hallmark of such environments is that while the auditory and visual temporal patterns are not necessarily correlated, there is some expectation that they match in their complexity. For deterministic sequential information, individuals are able to track two independent information streams without observable behavioral costs (e.g., Mayr, 1996). Neuroimaging studies show that individuals spontaneously track statistical features of different stimulus dimensions (e.g., category and location or shape and color; Aizenstein et al., 2004; Davis & Hasson, 2016). It is unclear, however, whether there are brain systems that holistically integrate statistical information across different modalities when exposed to multimodal inputs. This would be expected of brain regions involved in constructing a higher-level model of the environment.

In the current fMRI study we examined this issue by determining whether there are brain systems that signal a match between the levels of regularity in concurrently presented auditory and visual streams. We presented participants with four types of audiovisual stimuli that were constructed by independently manipulating the regularity of an auditory stream and a visual stream, which were jointly presented (see Table 1 for schematic of design). High-regularity auditory or visual streams allowed predictions about future events, whereas low-regularity streams were constructed to not allow predictions. Consequently, when crossing the two factors, the levels of regularity matched in two conditions: AudLow/VisLow and AudHigh/VisHigh. By contrast, in the two other conditions – AudLow/VisHigh and AudHigh/VisLow – the levels of regularity mismatched. In all conditions, the auditory and visual streams were *not* mutually informative. In other words, it was not possible to predict events in one stream from those in the other.

This design allowed us to identify brain areas showing an interaction between the levels of regularity in the auditory and visual streams and evaluate whether it reflects an effect of statistical mismatch effect, which we term “Mismatch in Multisensory Regularity” (MMR). This design further allowed us to identify brain regions that, in these multisensory contexts, showed sensitivity to the level of regularity of one stream independent of the level of regularity in the other (i.e., sensitivity to regularity of the auditory stream independently of that of the visual and vice versa).

Table 1. The four conditions in the study and graphical representation of a Mismatch in Multisensory Regularity (MMR) pattern with darker shades indicating greater BOLD signal change.

	Visual Regularity	
Auditory Regularity	Low	High
Low	AudLow/VisLow	AudLow/VisHigh
High	AudHigh/VisLow	AudHigh/VisHigh

We hypothesized that MMR signatures would be evident in two regions: first, the dorsal anterior cingulate cortex (ACC), which has been previously implicated statistical learning and signaling of match/mismatch effects and prediction error. Second, we expected to find signatures of multi-sensory regularity matching in regions implicated in integrating temporally unfolding low-level *sensory* features of auditory and visual stimuli (rather than ones based on distributional features). These include V5/MT+ (Sadaghiani, Maier, & Noppeney, 2009), the posterior superior temporal sulcus (pSTS; Tyll et al., 2013), superior temporal gyri bilaterally (Baumann & Greenlee, 2007), and visual area V3 (Ogawa & Macaluso, 2013). It has also been shown that when people observe movement sequences, matching audiovisual stimuli evoke greater activity in V5/MT+ than mismatching ones (Scheef et al., 2009). A similar sensitivity to an audiovisual match/mismatch has also been shown in studies that crossed visual and auditory motion effects (Alink, Singer, & Muckli, 2008; Rohe & Noppeney 2016; see also Soto-Faraco & Valjamae, 2012, for review).

Our second goal was to determine whether the magnitude of the MMR effect – formally an interaction term – depends on how strongly different individuals weigh the regularity of the auditory and visual streams. We hypothesized the MMR interaction term would be *weaker* for participants extremely sensitive to regularity in either the auditory or visual stream, as they might be focus on the regularity of one sensory stream in favor of the other. To this end, we first identified brain regions that, at the group level, most strongly tracked the regularity of the auditory stream independent of regularity of the visual stream (and vice versa). For each participant we then computed an effect-size for each of these two main effects. We similarly established the magnitude of the MMR effect for each participant from the region showing this interaction term at the group level, with higher values indicating stronger integration, where “integration” refers to the magnitude of the interaction term. This allowed us to then evaluate whether participants with higher sensitivity to within-modality regularity show weaker MMR effects.

Understanding whether there are brain regions that are sensitive to a match in the statistics of dual regularity streams is important both for neurobiological models of statistical learning, and multisensory integration. With respect to former, our current study fills a gap in current understanding, which has relied almost exclusively on studies of unisensory stimuli, and expands the scope of inquiry into the multisensory context. Furthermore, documenting a MMR signature would strongly inform a basic conundrum in the theoretical literature, which is whether there exist neural systems that are sensitive to regularity in multiple modalities (as reviewed in Frost, Armstrong, Siegelman, & Christiansen, 2015). With respect to theories of multisensory perception, our current study strongly bares on the question of whether there are brain systems that are sensitive to continuous, temporally extended features of multisensory streams. As discussed in prior work (Werner & Noppeney, 2011a), it appears that multisensory integration effects manifest most strongly in transient effects that occur on very short temporal scales, but largely do not show sustained cross-talk that is related temporally extended features such as the one manipulated here. Identifying MMR effects would suggest that the given region is sensitive to temporally extended features in both modalities.

2. Materials and methods

2.1 Participants

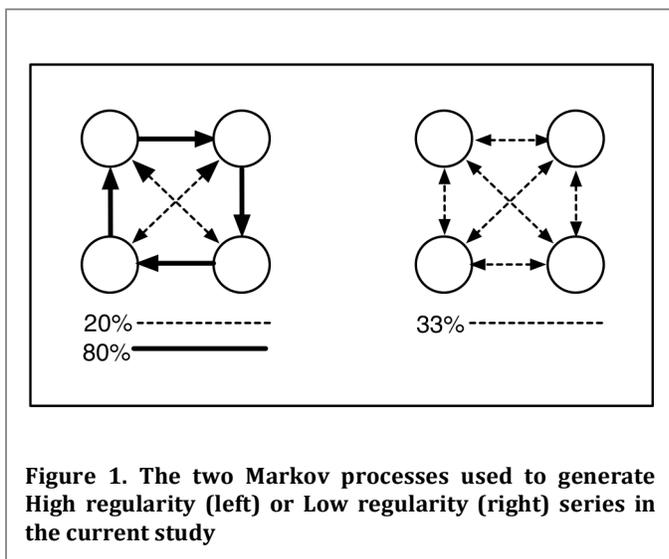
Nineteen right-handed participants (10 women, mean age = 22.63 years, SD = 3.37) from the University of Trento community participated in the study. None of them reported any history of neurological or psychiatric disturbance, visual or hearing impairment, or substance abuse. We excluded two participants' data due to excessive movement during the scan. A board-certified M.D. interviewed all candidate participants prior to participating in the study to evaluate for typical exclusion criteria. The Ethical Review Board of the University of Trento approved the study, and all expressed informed consent and were debriefed after the study.

2.2 Stimuli

We constructed audiovisual sequences consisting of four auditory tones and visually presented shapes in four locations. Auditory sequences consisted of pure tones at 261.63, 239.63, 392, and 440 Hz. The visual materials were shapes – a square, a circle, a triangle, and a star – colored red, blue, green, or yellow (in each series, the shape and color combinations were randomly assigned). Shapes appeared immediately above, below, left, or right of a central fixation cross at an eccentricity of 2 degrees. Their location was determined by the sequence order for the given level of regularity. Auditory tones and visual shapes were presented for 250ms with 53ms silent break (presentation rate = 3.3Hz). Each series consisted of 100 tokens and presented over 30sec

Regularity of the auditory and visual sequences could be one of two levels: “High” or “Low.” We constructed these two levels of regularity from separate Markov processes that varied by their respective transition constraints between items in the sequences (Figure 1). In the high regularity conditions, from each node there was 80% probability of transitioning to a second node, 20% probability of transitioning to a third node and 0% of transitioning to a fourth (repetitions were not allowed). In the low regularity condition, from each node, there was 33% probability of transitioning to each of the other nodes

(repetitions were not allowed). In both cases, the marginal frequencies of the 4 states was 25% – that is, the processes differed only in their transition probabilities, but not in their marginal frequencies. To construct the auditory and visual items’ sequencing in the two regularity levels, we generated 10,000 Markov chains each for high and low regularity. We calculated the transition probabilities for each of the 10,000 chains and selected the 24 sequences for which the sum of the absolute differences between the actual transition matrix and the ideal transition matrix (Figure 1) was minimal. An example of a stimulus stream where the regularity of both auditory and visual streams was high is given in Supplementary Movie 1.



With the two modalities (auditory, visual) each having each two levels of regularity (high, low), we created four combinations: 1) both auditory and visual with high regularity (“AudHigh/VisHigh”), 2) auditory with high regularity and visual with low regularity (“AudHigh/VisLow”), 3) auditory with low regularity and visual with high regularity (“AudLow/VisHigh”), and 4) both auditory and visual sequenced with low regularity (“AudLow/VisLow”). Note that in all cases, the visual and auditory transitions reflected different transition progressions so that there was no mutual information between the auditory and visual tracks, in any of the four conditions.

2.3 Procedure

Auditory stimuli were presented through headphones, and visual stimuli displayed via a mirror in the scanner. We presented the stimuli over four fMRI runs, each 525 seconds long. Each run included three presentations of each of the four conditions (i.e., 12 presentations per condition in total, over the entire study). We randomized the presentation order of the stimuli and catch trials throughout a run, such that it was unique for each run. We instructed participants to attend to the audio and visual stimuli and to press a button placed under their right index finger when presented with a catch trial. A catch trial presented a smiley face in the middle of the screen. Four catch trials (1.5 s each) were included in each run.

2.4 Image Acquisition

We acquired images using a 4T Bruker/Siemens system. Functional images were acquired with a single shot echo planar imaging blood oxygen level dependent (BOLD) sequence: 25 interleaved slices parallel to the AC/PC, TR = 1500 ms, TE = 33 ms, flip angle = 75 degrees, voxel size = 4 x 4 x 4.8 mm³, matrix = 64 x 64. Per run of 525 seconds, we collected 350 EPI scans. To acquire anatomical images, we used a 3D T1-weighted MPRAGE sequence. We collected 176 sagittal slices, with TR = 2700 ms, TE = 4ms, flip angle = 7 degrees, matrix size = 256 x 224, and isotropic voxel size = 1 x 1 x 1 mm³.

2.5 Data Analysis

2.5.1 Individual preprocessing

We processed each participant's anatomical images by implementing the following pipeline via FSL's *fsl_anat* script: Reorientation of the images to a standard space (MNI), cropping them around the circumference of the head to remove empty space in the images, and bias-field correction. The images were registered to a standard template by linear registration using FSL's *flirt* (Jenkinson and Smith, 2001; Jenkinson et al., 2002) and non-linear registration (FSL's *fnirt*; Andersson et al., 2007). This included FNIRT-based brain extraction and segmentation of the tissue-types using FSL's *fast* (Zhang et al., 2001). This segmentation allowed us to derive masks (used later as described below) for the white and gray matter and cerebrospinal fluid. We also segmented subcortical structures via FSL's *first*. In summary, for each participant's anatomy, we generated bias-field corrected anatomical images that were registered to standard space and for which we derived segmented images for each the white and gray matter, as well as subcortical structures.

We used AFNI (Cox, 1996) to process each participant's functional data as follows for each run. We discarded the first seven images to keep data collected after stabilization of the BOLD response. To attenuate the effect of spurious signal spikes in subsequent analyses, we de-spiked the time series. For each of the de-spiked time series we normalized values by the global mode. To increase the temporal signal-to-noise ratio of each voxel's time series, we applied a 6 mm spatial smoothing. We censored images by the DVARS method, which measures the rate of change in brain image intensity from one time point to the next (Power et al., 2012), using a criteria of 75th percentile + 1.5 times the interquartile range. We also regressed variance associated with signals of no interest ("nuisance signals") from the time series. These included regressors for signal in the white matter and cerebrospinal fluid areas, as well as motion parameters estimated by the displacement from the first image of the run. These regressions produced "clean" time series that constituted the signal of interest for further analyses. We aligned the time series of the first, second, and fourth functional runs to the first image of the third run using FSL's linear registration tool (*flirt*). This allowed us to model effects across the experiment for each participant instead of evaluating each run in isolation. For group analyses, to spatially transform individual space maps into group (MNI) space, we used boundary-based registration (Greve and Fischl, 2009) implemented in FSL to align the individual maps to group space via registrations of the bias-corrected anatomy to the first image of the third functional run for each individual.

2.5.2 Assessing regularity by modality and its interaction

For each participant, we fit the time series using generalized least squares with maximum likelihood (REML) estimation, with 4 regressors of interest reflecting the four conditions in the study (see section *Stimuli*) at every voxel in the brain. We used AFNI's 3dREMLfit software for this purpose, which implements a regression model that considers the serial autocorrelation in each voxel's time series (by also fitting an autoregressive model with 2 parameters), and thus produces a more accurate estimate of the statistical significance of regression terms on the single voxel level. This is important because we implemented a group-level analysis that takes into consideration not only the regression coefficient estimate, but also the T-value associated with each coefficient, as described below. Nuisance regressors included white matter and ventricle mask seeds.

On the single participant level for each voxel we constructed contrast terms to obtain a main effect for i) auditory regularity, (contrasting the two conditions in which auditory regularity was high versus those where it was low), ii) visual regularity, and iii) their interaction. We then used the resulting individuals' coefficients and t-statistics in a second-level group analysis.

For the group analysis, we modeled both within- and across-subject variability using AFNI's mixed effects meta-analysis program (Chen et al., 2012). We thresholded our results at the voxel-level by $p < .01$ (two-tailed) and required clusters that reached a volume of 1696 mm^3 to correct for family-wise error at $p < .05$. We determined this cluster correction volume by estimating the probability of false-positives in cluster sizes given the effective resolution, taking into account inherent spatial smoothing in our data. We estimated our data's resolution (estimation = $8.05 \times 8.24 \times 8.25 \text{ mm}^3$) from the full-width at half-maximum estimated from the residuals yielded in the mixed-effects meta-analysis.

In addition, we tested for sensitivity to uncertainty in anatomical regions previously linked to statistical learning for auditory or visual stimuli: left and right caudate, left and right putamen, and the left and right hippocampus. The putamen and caudate are thought to play a role in the construction of prediction and calculation of prediction error (e.g., Haruno & Kawato, 2006), though their role in prediction of auditory inputs is less clear (Hasson & Tremblay, 2016; Turk-Browne et al., 2010). The hippocampus has been suggested to be an important mediator of statistical learning, given its involvement in associative learning (e.g., Turk-Browne et al., 2010), and demonstrations of its sensitivity to input uncertainty (e.g., Harrison, Duggins, & Friston, 2006; Strange et al., 2005). However, our prior work that manipulated statistical regularities in unisensory streams (see Hasson, 2017 for review) has failed to identify hippocampal sensitivity to input uncertainty. From the subcortical segmentation described above, we used anatomical masks to derive an average signal over each region for each participant. We tested each of these regions for effects across participants using ANOVAs for repeated measures.

2.5.3 Modeling shorter durations for regularity by modality and its interaction

Our main analysis described above evaluated activation patterns over the entire 30 s presentation period of each series. We considered however that the effects of interest could have manifested earlier in the series and then decayed (in prior work we had documented effects in unisensory streams within 4-10sec; Nastase et al., 2014). We therefore also analyzed the data by restricting the block length analysis to the first 10, 15 or 20sec. This was implemented by excluding (censoring) from the regression model those data collected during the last 20 s, 15 s or 10 s of the series (respectively). Otherwise, the analyses were

identical to those described in the previous section. We implemented this only for the interaction term, because for the main effects we did not find results for some of the shorter block lengths.

To compare the results obtained for these different windows, from the group-level results we calculated the overlap of the group-level findings under each block length. We used the Sørensen-Dice similarity coefficient (Dice, 1945; Sørensen, 1948) for establishing similarity over sets. This similarity coefficient quantifies the product of twice the sets' intersection divided by the sum of occurrences in the sets. We calculated these similarity values for all significant voxels between successive block lengths, i.e., between significant voxels for 30 s and 20 s blocks, 20 s and 15 s blocks, and 15 s and 10 s blocks.

We also used a separate approach (Aizenstein et al., 2004) to differentiate between the earlier *process*-of-learning where statistical structure is extracted, from the later *result*-of-learning state, which is typified by sustained activity after an initial learning stage. Following Aizenstein et al. (2004) we defined a regressor for the 'process of learning' (POL) characterized by a decline in activity that asymptotes (schematically: 5,4,3,2,1,0,0,0,0,0) and a separate regressor for the 'result of learning' (ROL) characterized by a gradual increase that then remains at a high level (1,2,3,4,5,5,5,5,5,5). Aizenstein et al. (2004) showed it was possible to identify brain regions linked to these two sub-processes while participants were exposed to a very simple Markov process using visual stimuli.

These two regressors were convolved with a Hemodynamic response function. We then modeled these two separate processes for each of our 4 conditions, with the ascending/descending ramps modeled over 12 sec. In this way, we derived PoL and RoL beta estimates for each of the 4 conditions in the study. We then repeated the group-level analyses exactly as we had performed the main analysis, but separately for the PoL and RoL regressors.

2.5.4 Individual difference analysis

From the group-level analysis we selected peak voxels from brain regions that tracked regularity of the auditory stream or visual streams. We identified the peak voxel in these clusters and for each person derived the magnitude of the main effect in that voxel. This produced an Effect-Size Index for the auditory modality and another for the visual modality. In a similar manner we obtained the magnitude of the interaction term from peak voxels showing this pattern. We then computed pair-wise correlations between these indices. This established whether sensitivity to regularity in auditory streams correlated with sensitivity to regularity in visual streams, and whether sensitivity to regularity in either of the auditory or visual streams correlated with the magnitude of the interaction term.

3. Results

3.1 Modality-specific regularity tracking and mismatch effects

We identified two clusters that were sensitive to the level of auditory regularity in the audiovisual series (see Figure 2). One cluster (Figure 2B), straddling left anterior superior frontal gyrus (SFG), sulcus (SFS), and middle frontal gyrus (MFG), showed stronger deactivation for low-regularity auditory series than high-regularity ones. Another cluster (Figure 2A), around the right post-central gyrus, showed greater activity for low auditory regularity (see Figure 2C for activation levels).

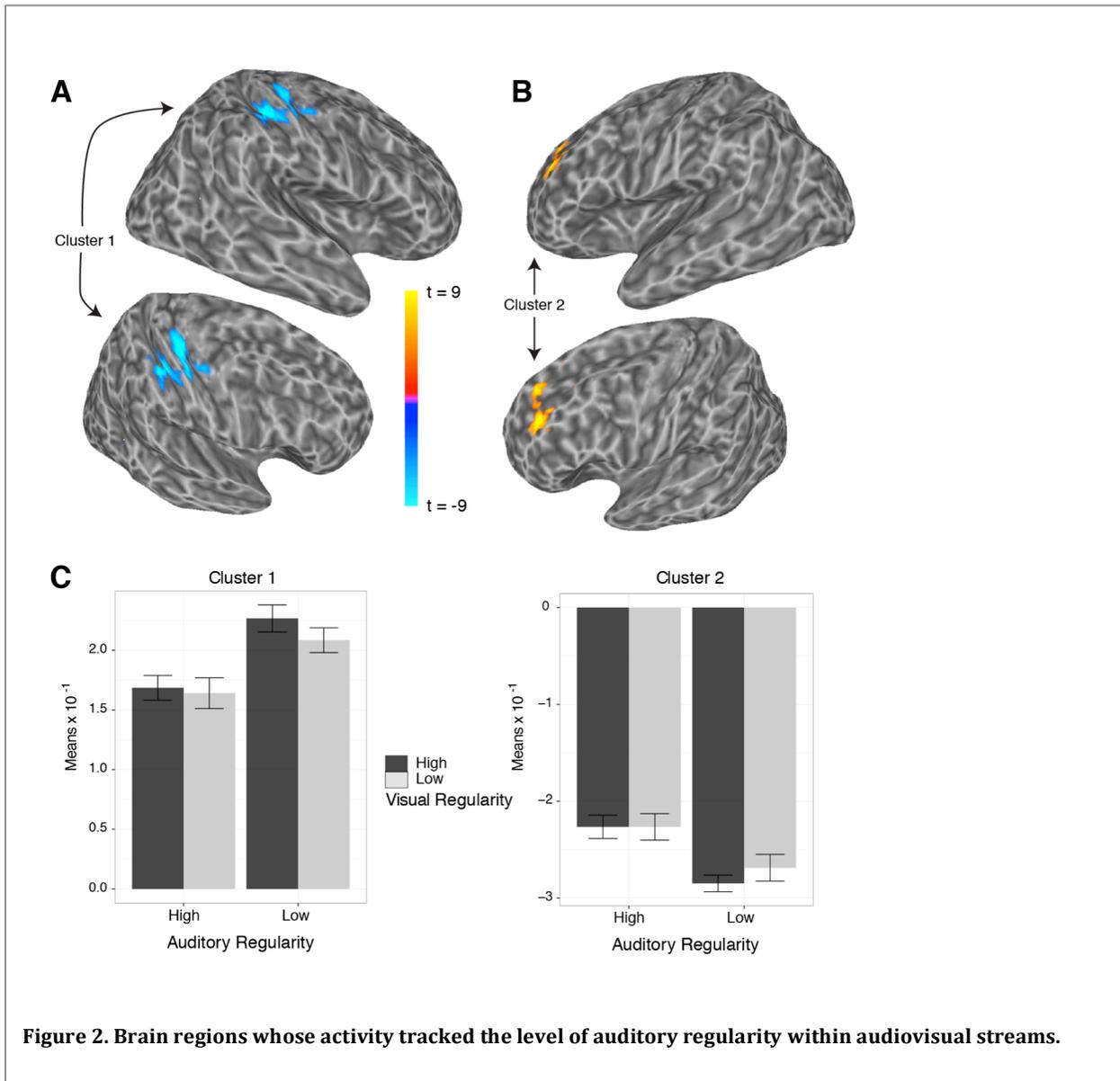
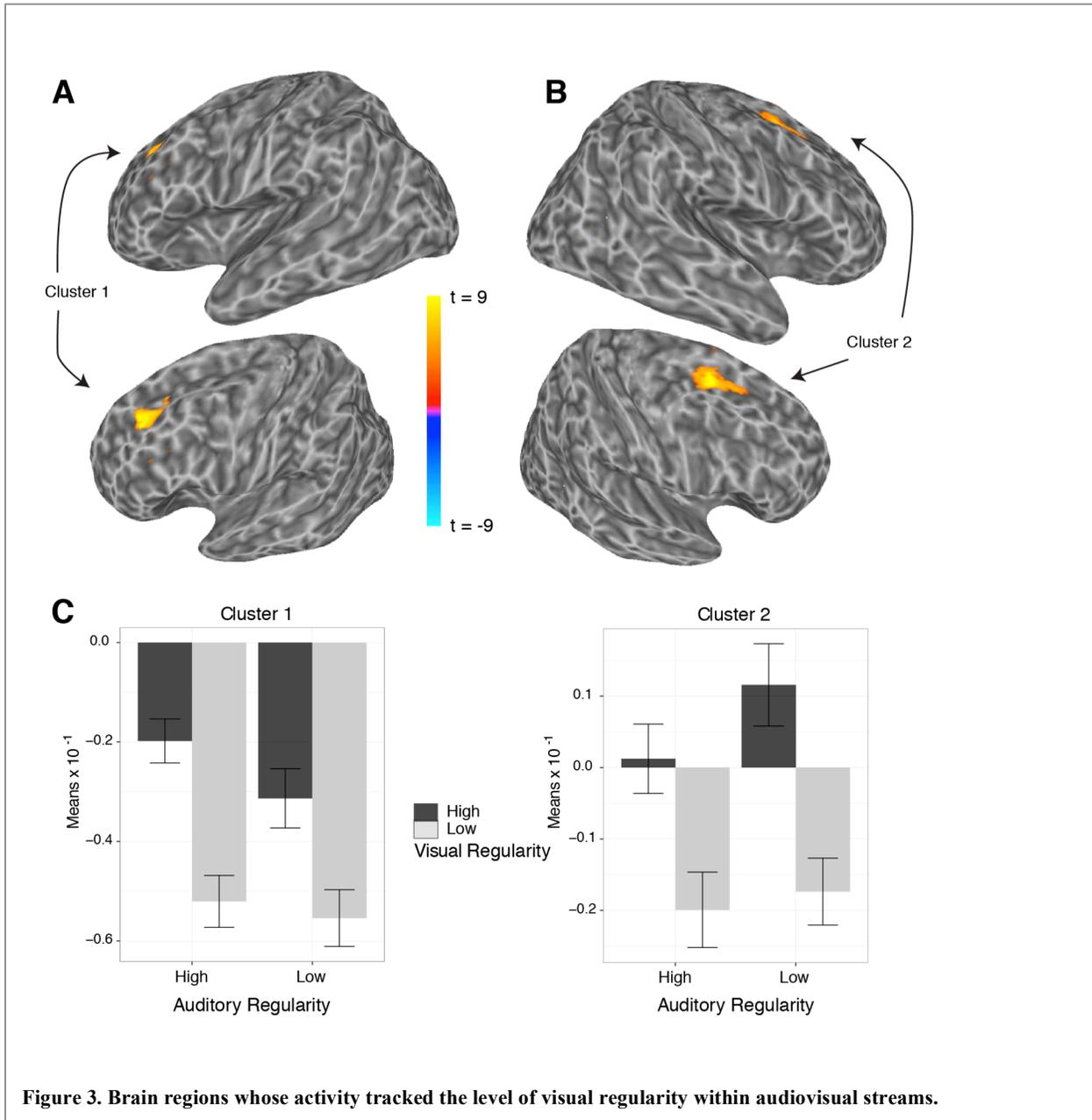


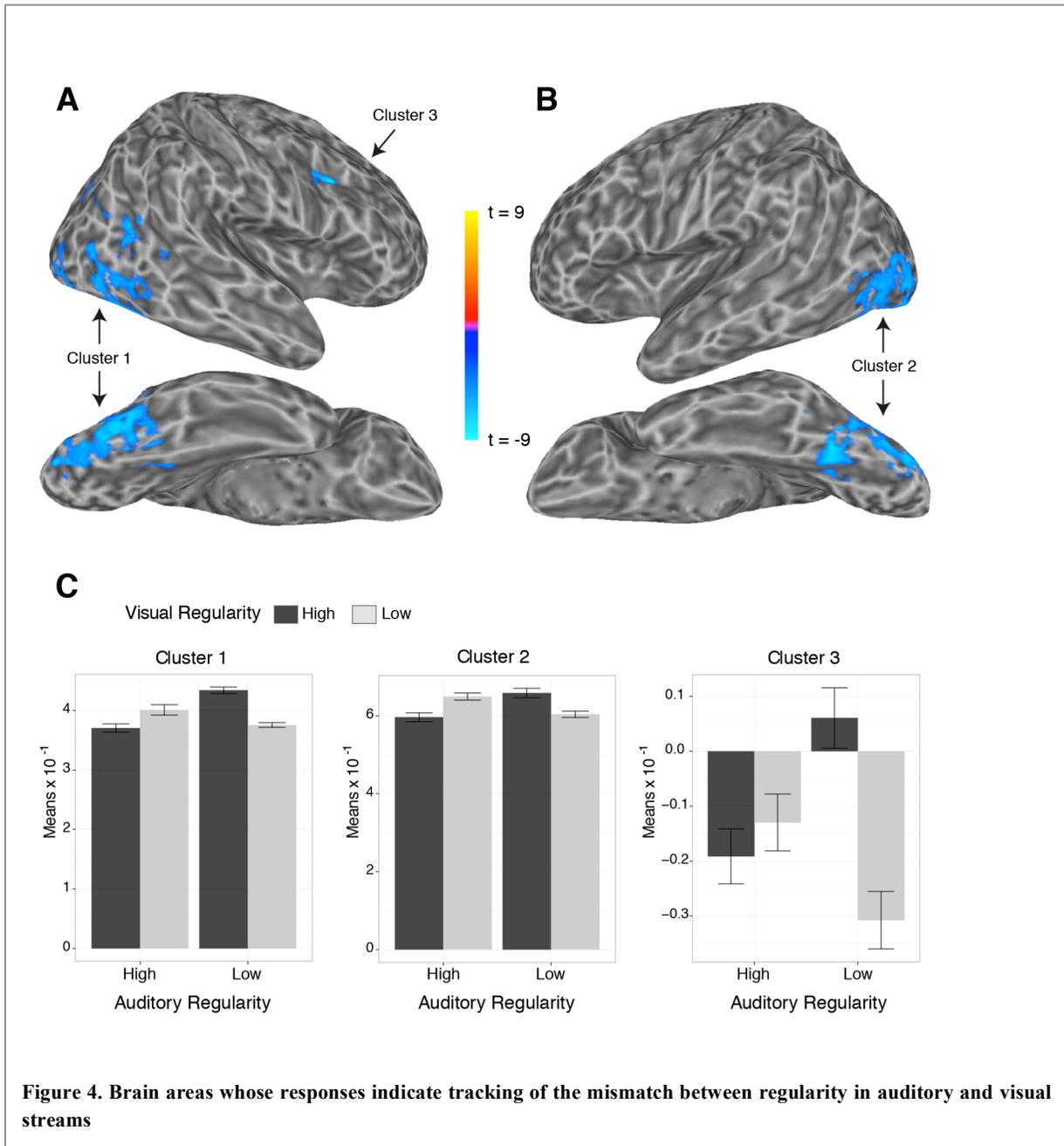
Figure 2. Brain regions whose activity tracked the level of auditory regularity within audiovisual streams.

We also found two clusters sensitive to visual regularity in the audiovisual series. In both clusters, higher regularity associated with higher activity. These clusters were found in the areas of left MFG (Figure 3A) and right MFG (Figure 3B, see Figure 3C for activation levels).



We found three clusters where BOLD responses during the series indicated sensitivity to the match between regularity of auditory and visual series (see Figure 4). As detailed below, in all cases, we found increased activity for the two conditions where the regularity of the auditory and visual series mismatched (AudHigh/VisLow; AudLow/VisHigh) compared to when they matched (AudHigh/VisHigh, AudLow/VisLow).

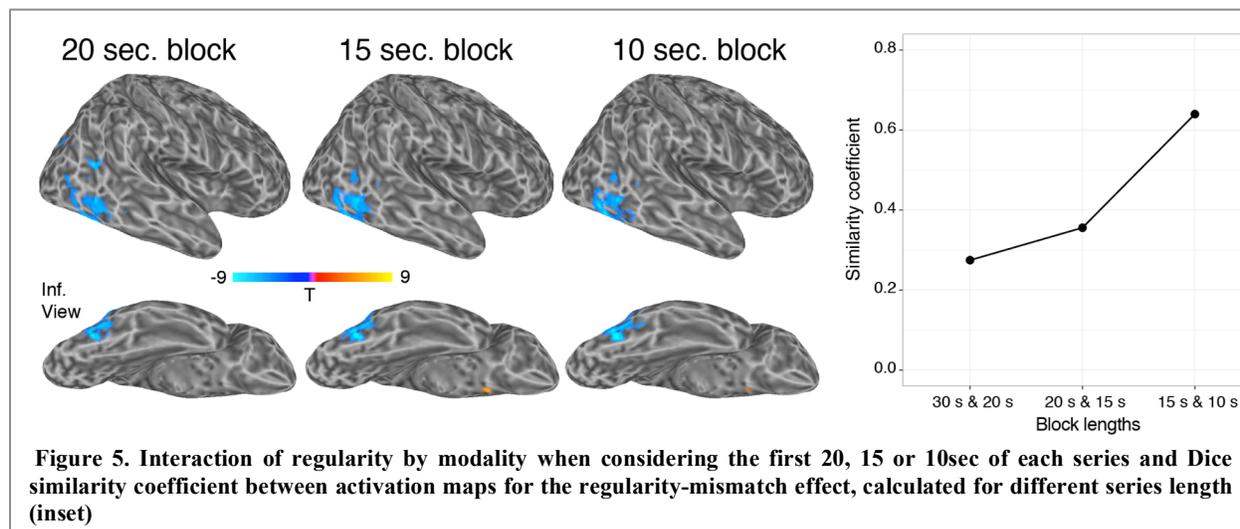
In the right hemisphere, a posterior cluster included parts of the posterior middle temporal gyrus (MTG), inferior temporal gyrus and fusiform, as well as inferior lateral occipital cortex (Figure 4A). In the left hemisphere, we found a posterior cluster in similar areas to those on the right, including the posterior inferior temporal cortex and fusiform (Figure 4B). In comparison to the right, the left hemisphere cluster centered more posteriorly, away from temporal cortices, covering much of middle and inferior occipital cortex. Finally, we found a small frontal cluster on the right, in the junction of right inferior frontal sulcus (IFS) and precentral gyrus (PCG). For this latter cluster, visual regularity appeared to be more strongly differentiated when auditory regularity was low than when it was high.



We also analyzed several anatomical regions of interest. However, none of these regions showed statistically significant main effects or interactions.

3.2 Analysis of shorter epochs

Although our series were 30 sec long, we examined whether the main results – the interaction of interest – developed earlier on by limiting our analyses to the first 10, 15 or 20sec of each series (see Methods). We found that in the right hemisphere, the interaction between auditory and visual modality developed within 10 s from the onset of the series (Figure 5). For the left hemisphere, we did not find an effect in these shorter blocks.



In comparing the overlap among significant group-level results for these different block-lengths, we found the strongest similarity for effects estimated from 10sec and 15sec blocks (Figure 5B). For the 30 s and 20 s blocks, the DICE coefficient was about 0.27. For the 20 s and 15 s blocks it increased, to 0.36, and for the 15 s and 10 s blocks it rose to 0.64. This suggests that between 15 and 20sec from stimulus appearance there was a shift in the activity structure that tracked the regularity match between the streams.

In addition, we implemented a separate approach for quantifying the learning dynamics (following Aizenstein et al., 2004), modeling for each condition an activity profile corresponding to a process-of-learning (POL; rapidly declining to baseline) and an activity profile corresponding to a result-of-learning (ROL; rapid ascent and sustained activity; see *Methods*). The results are presented in Figure 6. For POL we found that posterior left lateral temporal areas were sensitive to *auditory* regularity while the calcarine sulcus (V1) and two fronto-parietal regions were sensitive to *visual* regularity during POL. These findings are consistent with rapid detection of regularity for the auditory or visual streams, in multimodal contexts. For ROL we found that anterior and posterior midline regions (ACC, PCC) were sensitive to auditory regularity, and anterior midline regions were sensitive to visual regularity. This is consistent with differential levels of sustained attention as function of regularity later in the processing of the series.

Most interesting, the thalamus bilaterally and left intra-parietal sulcus showed an interaction term for POL, with the same pattern we reported previously: increased activity when the regularity of the two streams mismatched (Figure 7). This strongly suggests that that match/mismatch processes manifest during a rapid learning process in these regions.

[Figure 6 around here]

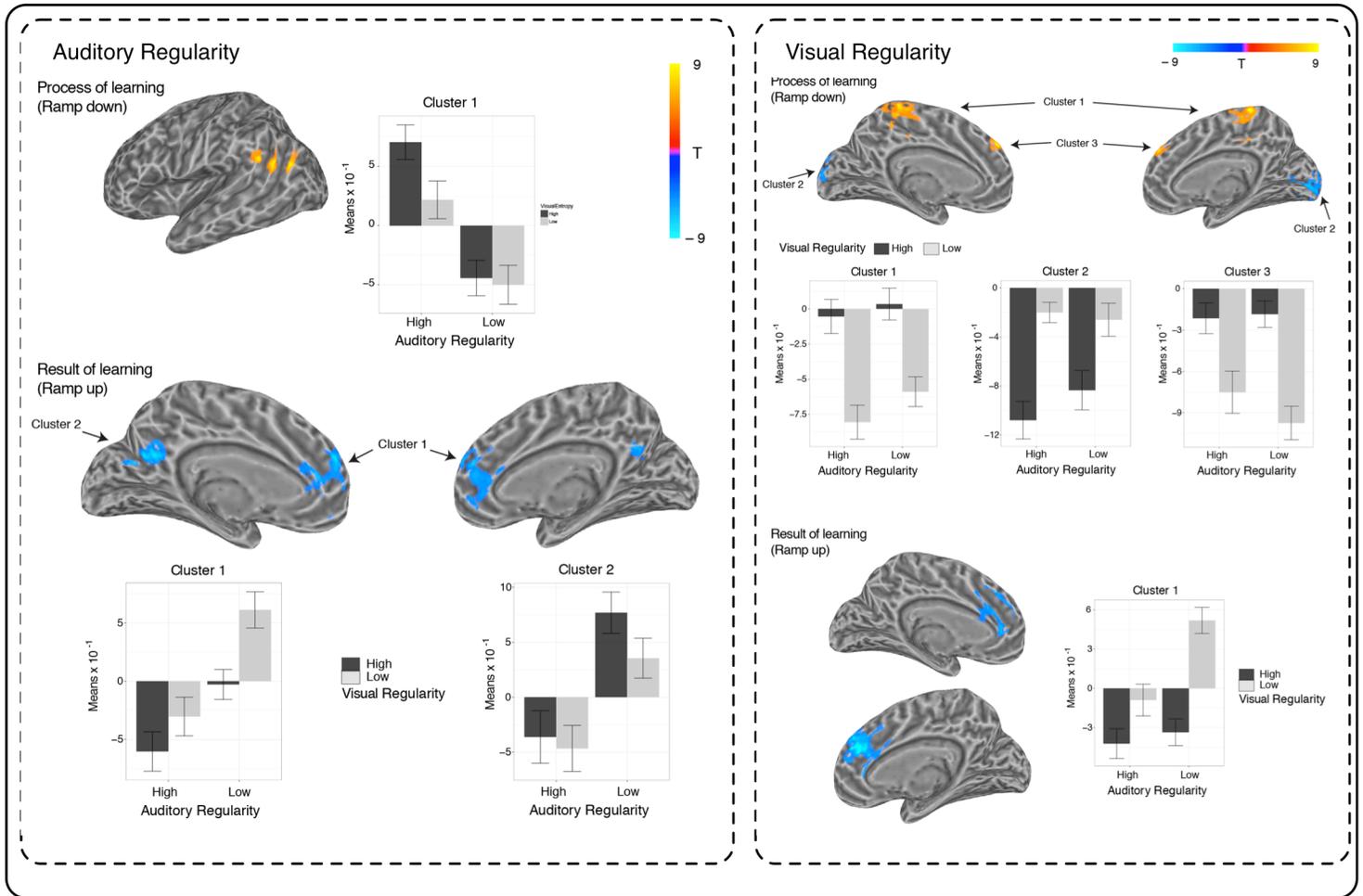
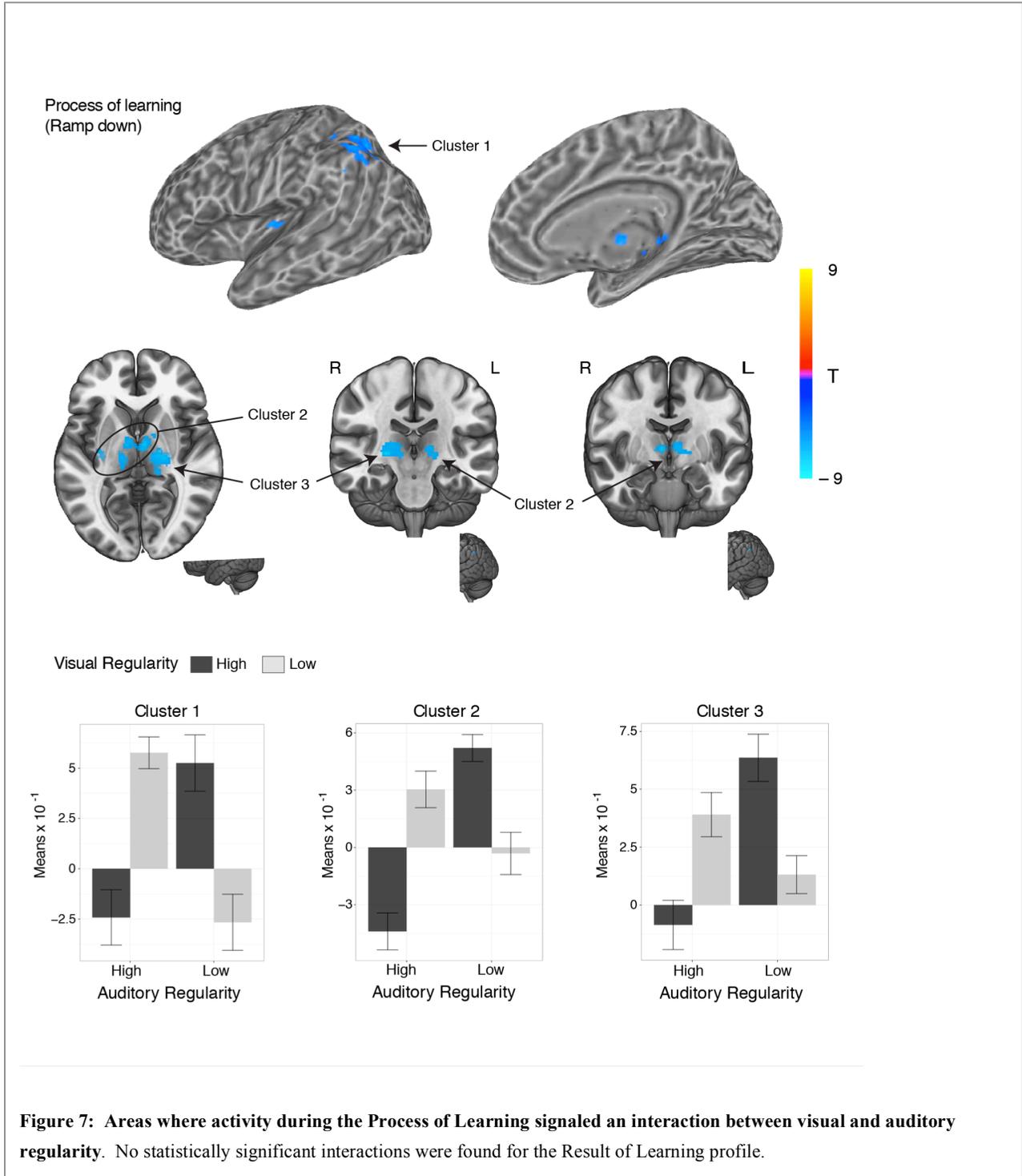


Figure 6: Areas that showed sensitivity to either auditory or visual regularity, tracking either a Process-of-learning or a Result-of-Learning profile.

[Figure 7 around here]



3.3 Stronger modality effects associate with weaker interaction terms

In this final analysis, we evaluated whether participants that showed a stronger sensitivity to auditory regularity also showed stronger sensitivity to visual regularity, or a stronger MMR interaction term. Focusing on the peak voxels in clusters 1, 2 for the auditory effect, clusters 1, 2 for the visual effect and clusters 1-3 for the interaction, we obtained the Beta value for the main effects and interaction per participant, and evaluated correlations between these values.

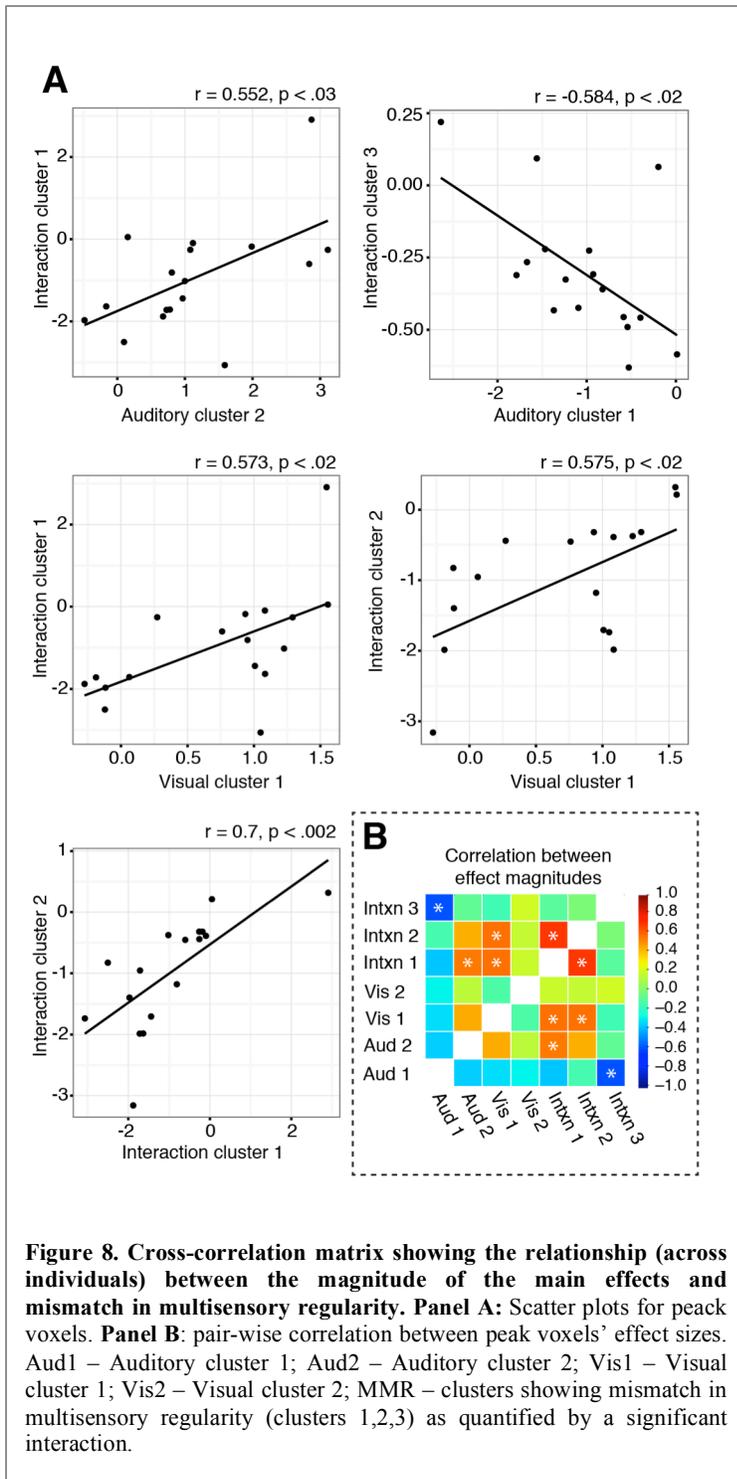


Figure 8 shows the resulting cross-correlation matrix and scatterplots for the significant correlations. Increased sensitivity to auditory regularity did not significantly correlate with increased sensitivity to visual regularity. However, we found that a reduced interaction term was correlated with increased

sensitivity to regularity in either the auditory or visual streams. Specifically, the scatter plots show that participants highly sensitive to auditory regularity or visual regularity showed practically no interaction term. Conversely, strong interaction terms accompanied a lack of sensitivity to auditory or visual regularity in the clusters where statistically significant main effects were found at the group level. Finally, the correlation between the magnitude of the interaction terms in the left and right occipital regions reached significance. Given the novelty of the analysis, we did not control these analyses for multiple comparisons. When implemented (False discovery rate), only the correlation between the interaction term magnitudes in interaction clusters 1 and 2 survived significance.

4. Discussion

Our findings demonstrate that during perception of multisensory audiovisual stimuli, there are brain systems that signal a mismatch between the level of regularity in the auditory and visual streams. This pattern is most evident in low and higher level visual systems, extending to V4 and V5 of the visual stream. Although these findings manifest on the group level, a detailed individual differences analysis suggests that the participants who show the strongest match/mismatch effects are those that show the weakest (neurobiological) signatures for tracking regularity in auditory or visual domains. These findings carry important implications for neurobiological theories of coding uncertainty and statistical learning, neurobiological theories of multisensory integration, and for functional accounts of individual differences in sensitivity to statistical structure in auditory and visual domains.

4.1 Sensitivity to regularity in multimodal contexts

To our knowledge, prior work has not examined the neurobiological substrates (or computational mechanisms) that mediate a combined influence of statistical information across simultaneously presented auditory and visual streams. The most strongly related behavioral work examined performance on serial reaction time (SRT) tasks, though for deterministic rather than stochastic sequences. That work suggested that coding of sequential information presented within separate stimulus dimensions comes at no cost (Mayr, 1996), with some suggesting that sequence learning in separate dimensions is essentially modular (Goschke & Bolte, 2012).

An initial neuroimaging study by Aizenstein et al. (2004) examined coding of multiple statistical dimensions in stochastic contexts and showed that individuals can track statistical regularities that govern separate dimensions of a stimulus stream (e.g., shape vs. color of simple shapes). Later work (Davis & Hasson, 2016) corroborated this conclusion but further suggested that processing regularity in independent dimensions is not modular (where “processing” holistically refers to the updating of statistical information, generation of predictions and prediction-error terms). Specifically, that work examined neural responses to stimuli where, over a stimulus series, the location of a subsequent visual image or its semantic category were either predictable or not. Several brain regions responded less optimally when both location and category were predictable, as contrasted to the conditions where only location or category was predictable. Thus, while it is possible to track regularities in separate dimensions, this might not be completed in an informationally encapsulated manner.

The current study extended these prior studies in that the stimuli we used could be tracked by separate sensory systems. Prior work that studied sensitivity to regularity or complexity in *uni-modal* auditory or

visual streams produced inconsistent findings on whether there are brain systems sensitive to regularity manipulations in a modality-independent manner. Nastase et al. (2014) identified multiple areas sensitive to manipulation of uncertainty in auditory or visual streams, but a conjunction of those only produced marginally significant results, involving left ventral premotor and central cingulate cortex. Another study (Schubotz & von Cramon, 2002) linked the ventral premotor cortex (bilaterally) to tracking predictability in auditory and visual series, though in slightly different locations.

In light of these prior findings, our results showing a ‘Mismatch in Multisensory Regularity’ (MMR) effect in the vicinity of right premotor cortex supports this region’s purported role in performing general computations related to regularity. In particular, the MMR effect is consistent with Wuerger et al.’s results (2012), which documented an inconsistency effect between auditory and visual (biological motion) information in this region, with greater activity for inconsistent multisensory information. Importantly, our findings indicate such mismatch effects in premotor cortex do not depend on visual information related to processing human body motion or actions more generally.

The MMR effects we document in visual cortex shed new light on the role of this region in coding for statistical regularity and expand current theories on the level of multisensory computations carried out in the visual processing stream. Most generally, it is not surprising that temporally-extended features of auditory stimuli impact activity in visual cortex; this has been shown both by studies examining visual cortex responses to uni-modal auditory streams and to multisensory audiovisual streams. To illustrate, a recent study used multivariate pattern analysis to determine whether human MT+ or LOC (defined independently from separate localizers) were sensitive to the directionality of auditory motion (Jiang, Stecker, & Fine, 2014). For typical participants (group size $n=7$) auditory direction could be decoded from the pattern of right LOC responses but not hMT+. However, another study that examined a similar question did successfully classify the direction of auditory motion from hMT+ activity (Dormal, Rezk, Yakobov, Lepore, & Collignon, 2016), adding to several other studies that documented audiovisual interactions in this region (von Salder & Noppeney, 2013). The involvement of right LOC in coding for auditory-motion direction was also indicated by Alink et al. (2012) via a whole-brain searchlight-based pattern classification.

Neuroimaging studies that specifically examined match/mismatch effects between auditory and visual movement have identified area V3 as a candidate area for integrating multisensory information (Ogawa & Macaluso, 2013). Another study found that area hMT/V5+ responded more strongly to coherent than incoherent visual and auditory spatial trajectories (Alink et al., 2008). An fMRI study of movie-segment viewing (de Haas, Schwarzkopf, Urner, & Rees, 2013) examined decoding accuracy for movie identity (movies were 3sec clips) based on activity in V1, V2 or V3. It found that decoding accuracy was lower when the visual and auditory tracks were incongruent (e.g., a movie of rooster crowing accompanied by ripping paper sound) than when they were congruent, or when there was no auditory track. (Yet a whole-brain searchlight analysis in that study did not reveal differences between the AV congruent and incongruent conditions). In MEG work, Jenkins et al. (Jenkins, Rhone, Idsardi, Simon, & Poeppel, 2011) examined the steady-state response profile to matching and mismatching auditory and visual signals. They found that response patterns around posterior temporal cortex indicated an interaction between the frequencies of the auditory and visual stimuli.

However, some of this aforementioned work (e.g., de Haas et al.) used short stimuli, leaving open the question of whether audiovisual integration also holds across stimuli that occur over longer temporal durations. Studies that specifically examined this issue suggested that, at least when using typical

movement-related manipulations, evidence for sustained audiovisual integration is weak. For instance, Giani et al. (2012), using MEG, did not find interactions between steady-state frequencies of auditory and visual sensory streams. In other words, there was no evidence for “true integration,” which would be manifested in power at the cross-modulation frequency. The authors therefore concluded, “information modulated in a steady fashion at different frequencies is integrated within but not across the senses” (p.1487). In another fMRI study, Werner and Noppeney (2011b) presented auditory and visual information (modulated at a rate of 0.1Hz) and modeled multisensory interaction effects at the onset of stimulation blocks, offset of blocks, and during the sustained (continuous) portion of the block. They found no evidence for multisensory interactions during the sustained portion of the block in any brain region. However, for right posterior STS they found stronger modulation at the 0.1Hz modulation rate in the audiovisual condition, suggesting this region integrates information, across modalities and over longer temporal scales.

In the current study, the strongest MMR effects implicated LOC bilaterally, extending into the fusiform. We also identified a more superior region on the right (falling within the same cluster), which may correspond to human V5/hMT+. However, given the variability of MT’s location across individuals, and the demonstrated reduced sensitivity of group-level results (rather than localizer-defined region-of-interest results), this remains suggestive (see (Dormal et al., 2016) for the importance of individually-defined localizers in addressing this question and evidence for coding of auditory motion direction in this region).

We also identified MMR effects in the thalamus bilaterally, as well as the left IPS, when specifically modeling a process-of-learning profile. The fact that these regions were *not* identified by the analysis that modeled the first 10sec via a sustained activation model (see Figure 5), suggests these regions may be sensitive to multisensory regularity in the most initial stages of processing. The IPS is a region often linked to anticipatory responses in cue-target paradigms (e.g., Egner et al., 2008) or in paradigms requiring future prediction based on stochastic statistical information (e.g., Huettel, Song, & McCarthy, 2005; Nastase et al., 2014).

It is important to keep in mind that the auditory-motion paradigm used in several prior studies may not form a very strong basis for understanding computations related to regularity (or uncertainty) effects. This is because auditory motion is ultimately a sensory feature of the stimulus (change in position). Studies that have specifically examined sensitivity to statistical regularity for auditory stimuli have provided limited understanding of the potential role of visual cortex in these computations. Yet one emerging finding is that activity in visual cortex tracks regularity of auditory streams. For instance, Tobia et al. (2012) found that activity in left lingual gyrus/calcarine, left fusiform gyrus and right fusiform gyrus tracked temporally unfolding features of simple auditory tonal streams (of the same type used in this study). Those regions’ BOLD time series during stimulus presentation correlated across individuals, indicating they were driven by exogenous features of the auditory stimuli (interestingly, this pattern was only found in visual regions). In another study (Tobia, Iacovella, Davis, et al., 2012), changes in visual cortex activity preceded subjective experiences of changes in statistical regularities of auditory streams. Altogether, the MMR effect we document in the occipital system is consistent with the theoretical position that these regions hold the capacity for integrating information across modalities in a sustained manner, that is, over longer temporal time frames (Werner and Noppeney, 2011).

4.2 Individual differences in responses to statistical regularities and their link to multisensory integration

Our analysis of individual differences in sensitivity to the regularity of the auditory and visual streams suggests that it is important to consider this sort of variation when accounting for sensitivity to regularity in multisensory contexts. To recap, within clusters found in the group-level analysis, we identified the peak-voxel showing the effect and extracted an index of effect size (the Beta value of the contrast) for each participant. We could then evaluate the relation between different effect sizes. Given that our study is the first to evaluate this question, we did not control for family-wise error across the six correlation tests, and so it is important to replicate these findings in future work. Nonetheless, as we discuss below, noticeably absent was a correlation between sensitivity to regularity in the auditory domain and sensitivity to regularity in the visual domain. In fact, their correlations were neither significantly positive nor negative. On the modular view of dimension coding (Mayr, 1996), it may be expected that individual differences in coding statistical regularity would reflect differences in attention. This posits that coding for structure in different dimensions operates in parallel, so more attentive participants would be sensitive to structure in both the auditory and visual sequence. We did not find this pattern. Recent behavioral research further suggests there might not be a strong common functional basis for the learning of structure (statistical learning) within auditory and visual streams. Siegelman and Frost (2015) reported negligible correlations between performance on visual and auditory statistical learning tasks (even though each of the tasks had reasonable test-retest correlations exceeding $r = 0.55$). They concluded that the capacity to discover structure in one sensory modality does not indicate a similar capacity for another modality.

We did, however, document correlations between sensitivity to regularity in the auditory and visual streams and the magnitude of the interaction contrast, estimated from each of the three clusters that demonstrated an interaction effect. Depicting these correlations in scatterplots allowed determining the absolute effect sizes (these are meaningful magnitudes, which are fundamental for interpreting the results). For the frontal cluster (Figure 4 cluster #3) we found that participants who showed the strongest interaction term in that cluster also most weakly discriminated auditory regularity, which we quantified from a peak voxel in right motor cortex (Figure 2 cluster #1). The magnitude of the interaction term in right visual cortex (Figure 4 cluster #2) correlated with sensitivity to visual regularity, which we measured in left frontal cortex (Figure 2 cluster #1). Here too, greater sensitivity to regularity (in the visual stream) associated with a reduced interaction term (approaching 0). The interactions for the left visual cortex (Figure 4 cluster #1) showed the same pattern. But here we found that stronger signatures for an interaction associated with both weaker sensitivity to auditory regularity (as measured in left frontal cortex, Figure 2 cluster #2) and weaker sensitivity to visual regularity (as measured in left frontal cortex, Figure 3 cluster #1).

Taken together, these findings suggest a balance, or trade-off, between coding of regularity in “sensory-specific” systems that are more strongly sensitive to regularity in either the auditory or visual streams and, conversely, those systems that are sensitive to the relation between regularity in auditory and visual streams.

4.3 Limitations and future directions

A limitation of our study is that it cannot determine exactly what participants were doing when processing the different multimodal series, and it is therefore unclear whether statistical learning in such multimodal

contexts reflects a completely new learning phenomenon as compared to unisensory contexts. As mentioned in the introduction, Mayr (1996) concluded that when independent streams are presented in parallel, their structure is apparently extracted without mutual interference. In addition, behavioral studies contrasting statistical learning for unisensory and multisensory inputs have not documented robust differences, and the data on this issue are contradictory. Seitz et al. (2007) examined statistical learning for audiovisual stimuli (we use the format 1/A to refer to a stimulus with auditory token ‘1’ and visual token ‘A’) that were presented as pairs, e.g., [1/A; 2/B]. In a subsequent test stage, participants were more accurate in recognizing the complete audiovisual transition (1/A-2/B) than in recognizing transitions between auditory stimuli alone [1-2] or visual stimuli alone [A-B]. However, another study (Glicksohn & Cohen, 2013) conversely suggested that learning audiovisual associations is more demanding. Participants were presented with triplets of auditory [1-2-3], visual [A-B-C] or audiovisual [1/A- 2/B-3/C] stimuli that formed fixed “words” presented sequentially amongst other such words. At the test stage, performance for audiovisual pairs was significantly reduced as compared to a baseline auditory condition that did not include visual stimuli. Yet, another study (Mitchel & Weiss, 2011) compared statistical learning performance for auditory, visual, or audiovisual streams (the latter composed of fixed audiovisual pairs as in our current study) but showed no reduction in performance in the audiovisual condition. Thus, prior behavioral work does not provide a basis for thinking that sensitivity to statistical features of multisensory streams relies on fundamentally different computations. That said, the regions we identify in the current study as showing an MMR effect, or sensitivity to the regularity of auditory or visual streams, are not ones typically implicated in prior uni-sensory studies. For example, here we identified the middle frontal gyrus as sensitive to regularity of visual streams, but prior work has typically implicated more extensive fronto-parietal or medial temporal systems (see Hasson, 2017 for review). A direct contrast between multisensory and unisensory conditions is needed to address this issue.

In addition, while we find signatures of multisensory integration in higher-level regions of the visual stream, in absence of a detailed, testable computational-neuroscientific model it is difficult to account for the mechanism that produces these interaction effects. One can consider at least two possible types of computations that may give rise to such effects. On the first, information about statistical regularities is transduced into a motion-like code in brain systems that track movement, and that may prefer consistency in codes derived from auditory and visual inputs. Another, related possibility is that the identified areas are part of a sensory ‘prediction-evaluation circuit’ where predictions about the statistical structure of one modality are made or evaluated on the basis of information in the other modality, thus generating cross-modal mismatch effects due to greater magnitude of error terms. This would entail a capacity to make relatively long term predictions about the future's statistical features. We note that both of these depart from typical accounts of multisensory integration that intend to explain for how sensory signals combine to optimize processing of single, discrete, distal events.

Finally, our study aimed to identify systems tracking regularity in multiple modalities by using a contemporaneous multimodal design where the two streams are presented synchronously. Another approach would be to present unisensory streams and determine whether any brain region sensitivity to regularity for both auditory and visual inputs, indicating supra-modal involvement (an issue we had examined previously; Nastase, Iacovella, & Hasson, 2014). These two theoretical approaches to multimodality/cross-modality are manifestly different, because brain systems demonstrating a conjunction pattern as defined above might still not show the MMR effects we identify in the current study. Directly examining these two issues could be an interesting avenue for future work. Another

potential limitation is that we had presented the two series in a time-locked manner, which could have driven attention to audio-visual combinations (even though the two streams were not correlated). Presenting the two streams out of phase could produce different outcomes.

4.4 Summary

We present findings that identify, for the first time, brain systems that are sensitive to the match in statistical regularity across multiple input streams. We find that lateral occipital cortex, bilaterally, and the right ventral premotor cortex show this pattern, with more specific signatures found in parietal cortex and thalamus during the early part of series presentation. In all cases, this was seen in a “regularity matching” response pattern, with stronger activity for stimuli where the regularity of auditory and visual streams mismatched. Still, other brain regions tracked the regularity of the auditory and visual streams. The fact that stronger signatures of regularity tracking in these latter regions was linked to weaker signatures of regularity matching in visual cortex suggests that multisensory regularity matching effects depend on balanced processing of both input modalities.

References

- Aizenstein, H. J., Stenger, V. A., Cochran, J., Clark, K., Johnson, M., Nebes, R. D., & Carter, C. S. (2004). Regional brain activation during concurrent implicit and explicit sequence learning. *Cereb Cortex, 14*(2), 199-208.
- Alink, A., Euler, F., Kriegeskorte, N., Singer, W., & Kohler, A. (2012). Auditory motion direction encoding in auditory cortex and high-level visual cortex. *Hum Brain Mapp, 33*(4), 969-978. doi: 10.1002/hbm.21263
- Alink, A., Singer, W., & Muckli, L. (2008). Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *J Neurosci, 28*(11), 2690-2697. doi: 10.1523/JNEUROSCI.2980-07.2008
- Baumann, O., & Greenlee, M. W. (2007). Neural correlates of coherent audiovisual motion perception. *Cereb Cortex, 17*(6), 1433-1443. doi: 10.1093/cercor/bhl055
- Bischoff-Grethe, A., Proper, S. M., Mao, H., Daniels, K. A., & Berns, G. S. (2000). Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *The Journal of Neuroscience, 20*(5), 1975-1981.
- Cunillera, T., Camara, E., Toro, J. M., Marco-Pallares, J., Sebastian-Galles, N., Ortiz, H., . . . Rodriguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage, 48*(3), 541-553. doi: 10.1016/j.neuroimage.2009.06.069
- Davis, B., & Hasson, U. (2016). Predictability of what or where reduces brain activity, but a bottleneck occurs when both are predictable. *Neuroimage*. doi: 10.1016/j.neuroimage.2016.06.001
- de Haas, B., Schwarzkopf, D. S., Urner, M., & Rees, G. (2013). Auditory modulation of visual stimulus encoding in human retinotopic cortex. *NeuroImage, 70*, 258-267. doi: 10.1016/j.neuroimage.2012.12.061
- Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron, 88*(1), 2-19. doi: 10.1016/j.neuron.2015.09.019
- Dormal, G., Rezk, M., Yakobov, E., Lepore, F., & Collignon, O. (2016). Auditory motion in the sighted and blind: Early visual deprivation triggers a large-scale imbalance between auditory and "visual" brain regions. *NeuroImage, 134*, 630-644. doi: 10.1016/j.neuroimage.2016.04.027
- Egner, T., Monti, J. M., Trittschuh, E. H., Wieneke, C. A., Hirsch, J., & Mesulam, M. M. (2008). Neural integration of top-down spatial and feature-based information in visual search. *J Neurosci, 28*(24), 6141-6151. doi: 10.1523/JNEUROSCI.1262-08.2008
- Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn Sci, 19*(3), 117-125. doi: 10.1016/j.tics.2014.12.010

- Giani, A. S., Ortiz, E., Belardinelli, P., Kleiner, M., Preissl, H., & Noppeney, U. (2012). Steady-state responses in MEG demonstrate information integration within but not across the auditory and visual senses. *NeuroImage*, *60*(2), 1478-1489. doi: 10.1016/j.neuroimage.2012.01.114
- Glicksohn, A., & Cohen, A. (2013). The role of cross-modal associations in statistical learning. *Psychon Bull Rev*, *20*(6), 1161-1169. doi: 10.3758/s13423-013-0458-4
- Goschke, T., & Bolte, A. (2012). On the modularity of implicit sequence learning: independent acquisition of spatial, symbolic, and manual sequences. *Cogn Psychol*, *65*(2), 284-320. doi: 10.1016/j.cogpsych.2012.04.002
- Harrison, L. M., Bestmann, S., Rosa, M. J., Penny, W., & Green, G. G. (2011). Time scales of representation in the human brain: weighing past information to predict future events. *Front Hum Neurosci*, *5*, 37. doi: 10.3389/fnhum.2011.00037
- Haruno, M., & Kawato, M. (2006). Different neural correlates of reward expectation and reward expectatio error in the putamen and caudate nucleus during stimulusaction- reward association learning. *J Neurophysiol*, *95*(2), 948-959. doi: 10.1152/jn.00382.2005
- Harrison, L. M., Duggins, A., & Friston, K. J. (2006). Encoding uncertainty in the hippocampus. *Neural Networks*, *19*(5), 535-546.
- Hasson, Uri, & Tremblay, Pascale. (2016). Neurobiology of Statistical Information Processing in the Auditory Domain. In Steven L. Small & G. Hickok (Eds.), *Neurobiology of Language* (pp. 527-537). San Diego: Academic Press.
- Huettel, S. A., Mack, P. B., & McCarthy, G. (2002). Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. *Nat Neurosci*, *5*(5), 485-490. doi: 10.1038/nn841
- Huettel, S. A., Song, A. W., & McCarthy, G. (2005). Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *J Neurosci*, *25*(13), 3304-3311. doi: 10.1523/JNEUROSCI.5070-04.2005
- Jenkins, J., 3rd, Rhone, A. E., Idsardi, W. J., Simon, J. Z., & Poeppel, D. (2011). The elicitation of audiovisual steady-state responses: multi-sensory signal congruity and phase effects. *Brain Topogr*, *24*(2), 134-148. doi: 10.1007/s10548-011-0174-1
- Jiang, F., Stecker, G. C., & Fine, I. (2014). Auditory motion processing after early blindness. *J Vis*, *14*(13), 4. doi: 10.1167/14.13.4
- Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain Lang*, *127*(1), 46-54. doi: 10.1016/j.bandl.2012.11.007
- Leonard, M. K., Bouchard, K. E., Tang, C., & Chang, E. F. (2015). Dynamic encoding of speech sequence probability in human temporal cortex. *J Neurosci*, *35*(18), 7203-7214. doi: 10.1523/JNEUROSCI.4100-14.2015
- Mayr, U. (1996). Spatial attention and implicit sequence learning: evidence for independent learning of spatial and nonspatial sequences. *J Exp Psychol Learn Mem Cogn*, *22*(2), 350-364.

- McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *The Journal of Neuroscience*, *26*(29), 7629-7639. doi: 10.1523/JNEUROSCI.5501-05.2006
- Mitchel, A. D., & Weiss, D. J. (2011). Learning across senses: cross-modal effects in multisensory statistical learning. *J Exp Psychol Learn Mem Cogn*, *37*(5), 1081-1091. doi: 10.1037/a0023700
- Nastase, S. A., Iacovella, V., & Hasson, U. (2014). Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Hum Brain Mapp*, *35*(4), 1111-1128. doi: 10.1002/hbm.22238
- Ogawa, A., & Macaluso, E. (2013). Audio-visual interactions for motion perception in depth modulate activity in visual area V3A. *Neuroimage*, *71*, 158-167. doi: 10.1016/j.neuroimage.2013.01.012
- Sadaghiani, S., Maier, J. X., & Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *J Neurosci*, *29*(20), 6490-6499. doi: 10.1523/JNEUROSCI.5437-08.2009
- Scheef, L., Boecker, H., Daamen, M., Fehse, U., Landsberg, M. W., Granath, D. O., . . . Effenberg, A. O. (2009). Multimodal motion processing in area V5/MT: evidence from an artificial class of audio-visual events. *Brain Res*, *1252*, 94-104. doi: 10.1016/j.brainres.2008.10.067
- Schubotz, R. I., & von Cramon, D. Y. (2002). Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. *Neuroimage*, *15*(4), 787-796. doi: 10.1006/nimg.2001.1043
- Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, *36*(10), 1445-1453.
- Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *J Mem Lang*, *81*, 105-120. doi: 10.1016/j.jml.2015.02.001
- Soto-Faraco, S., & Valjamae, A. (2012). Multisensory Interactions during Motion Perception: From Basic Principles to Media Applications. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. Boca Raton (FL).
- Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw*, *18*(3), 225-230. doi: S0893-6080(05)00006-7 [pii] 10.1016/j.neunet.2004.12.004
- Tobia, M. J., Iacovella, V., Davis, B., & Hasson, U. (2012). Neural systems mediating recognition of changes in statistical regularities. *Neuroimage*, *63*(3), 1730-1742. doi: 10.1016/j.neuroimage.2012.08.017
- Tobia, M. J., Iacovella, V., & Hasson, U. (2012). Multiple sensitivity profiles to diversity and transition structure in non-stationary input. *Neuroimage*, *60*(2), 991-1005. doi: 10.1016/j.neuroimage.2012.01.041
- Tremblay, P., Baroni, M., & Hasson, U. (2013). Processing of speech and non-speech sounds in the supratemporal plane: auditory input preference does not predict sensitivity to statistical structure. *Neuroimage*, *66*, 318-332. doi: 10.1016/j.neuroimage.2012.10.055

- Tremblay, P., Deschamps, I., Baroni, M., & Hasson, U. (2016). Neural sensitivity to syllable frequency and mutual information in speech perception and production. *Neuroimage*. doi: 10.1016/j.neuroimage.2016.05.018
- Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *The Journal of Neuroscience*, *30*(33), 11177-11187. doi: 10.1523/JNEUROSCI.0858-10.2010
- Tyll, S., Bonath, B., Schoenfeld, M. A., Heinze, H. J., Ohl, F. W., & Noesselt, T. (2013). Neural basis of multisensory looming signals. *Neuroimage*, *65*, 13-22. doi: 10.1016/j.neuroimage.2012.09.056
- von Saldern, S., & Noppeney, U. (2013). Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. *J Neurosci*, *33*(20), 8841-8849. doi: 10.1523/JNEUROSCI.3020-12.2013
- Werner, S., & Noppeney, U. (2011a). The contributions of transient and sustained response codes to audiovisual integration. *Cerebral Cortex*, *21*(4), 920-931.
- Werner, S., & Noppeney, U. (2011b). The contributions of transient and sustained response codes to audiovisual integration. *Cereb Cortex*, *21*(4), 920-931. doi: 10.1093/cercor/bhq161
- Wuerger, S. M., Parkes, L., Lewis, P. A., Crocker-Buque, A., Rutschmann, R., & Meyer, G. F. (2012). Premotor cortex is sensitive to auditory-visual congruence for biological motion. *J Cogn Neurosci*, *24*(3), 575-587. doi: 10.1162/jocn_a_00173