NeuroImage 157 (2017) 648–659

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

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

Michael Andric[a,*], Ben Davis[a], Uri Hasson[a,b]

[a] *Center for Mind/Brain Sciences (CIMeC), University of Trento, Italy*
[b] *Center for Practical Wisdom, The University of Chicago, United States*

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 suggested 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.

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 et al., 2011; Huettel et al., 2002; Strange et al., 2005; Turk-Browne et al., 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 et al., 2006; Nastase et al., 2014; Tobia et al., 2012b; Tremblay et al., 2013), with some implicating the basal ganglia (e.g., Geiser et al., 2012; McNealy et al., 2006). Some lateral-temporal regions also track sub-lexical statistics in natural language (e.g., Leonard et al., 2015; Tremblay et al., 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 and Frost, 2015). It also suggests that a domain-general capacity, if existent, would be subservient to modality-specific processing constraints (Frost et al., 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 spon-

---

[*] Corresponding author.
E-mail address: andric.michael@gmail.com (M. Andric).

http://dx.doi.org/10.1016/j.neuroimage.2017.05.028
Received 11 March 2017; Accepted 15 May 2017

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

taneously track statistical features of different stimulus dimensions (e.g., category and location or shape and color; Aizenstein et al., 2004; Davis and 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).

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 et al., 2009), the posterior superior temporal sulcus (pSTS; Tyll et al., 2013), superior temporal gyri bilaterally (Baumann and Greenlee, 2007), and visual area V3 (Ogawa and 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 et al., 2008; Rohe and Noppeney 2016; see also Soto-Faraco and 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 et al., 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 and Noppeney, 2011), 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.

Materials and methods

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.

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 250 ms with 53 ms silent break (presentation rate = 3.3 Hz). Each series consisted of 100 tokens and presented over 30 s.