Auditory Target Detection Enhances Visual Processing and Hippocampal Functional Connectivity

Abbreviated Title: Temporal Selection Boosts Visual Processing

Roy Moyal[1], Hamid B. Turker[1], Wen-Ming Luh[2], & Khena M. Swallow[1] *

1. Department of Psychology & Cognitive Science Program, Cornell University, Ithaca, NY, 14853

2. National Institute on Aging, National Institutes of Health, Baltimore, MD, 21225

*To whom correspondence should be addressed:

Khena Swallow (kms424@cornell.edu)

Number of pages: 23

Number of figures: 5

Number of tables: 2

Number of words:

Abstract: 250

Introduction: 643

Discussion: 1493

Conflict of Interest Statement: The authors declare no competing financial interests.

Acknowledgments: This work was supported by an NIH NCRR grant [1S10RR025145] to the Cornell Magnetic Resonance Imaging Facilities and by the College of Arts and Sciences at Cornell University. The authors would like to thank Suyash Bhogawar, Bohan Li, Alyssa Phelps, Roy Proper, and Emily Qualls for their assistance with data collection. We also thank Elizabeth Riley for her help with identification of the locus coeruleus, and Adam Anderson, Eve DeRosa, and Nathan Spreng for discussion of this work.

Abstract. Though dividing one's attention between two input streams typically impairs performance, detecting a behaviorally relevant stimulus can sometimes enhance the encoding of task-irrelevant information presented at the same time. Previous research has shown that temporal selection of this kind boosts visual cortical activity and incidental memory. An important and yet unanswered question is whether such effects are reflected in processing quality and functional connectivity in visual regions and the hippocampus. In this fMRI study, participants were asked to memorize a stream of images and press a button when they heard an auditory tone of a prespecified pitch. Images could be presented with a target tone, with a distractor tone, or without a tone. Auditory target detection increased activity throughout the ventral visual cortex but lowered it in the hippocampus. These effects were accompanied by a widespread enhancement in functional connectivity between the ventral visual cortex and the hippocampus. Image category classification accuracy was higher on target tone trials than on distractor and no tone trials in the fusiform gyrus and the parahippocampal gyrus. This effect was stronger in clusters whose activity was more correlated with the hippocampus on target tone than on distractor tone trials. In agreement with accounts suggesting that subcortical noradrenergic influences play a role in temporal selection, auditory target detection also caused an increase in locus coeruleus activity and phasic pupil responses. These findings outline a network of cortical and subcortical regions that are involved in the selection and processing of information presented at behaviorally relevant moments.

Significance Statement. Attention influences the degree to which we remember everyday experiences. This study examines the neural mechanisms involved in committing important events to memory. It links the selection of important information in time (temporal selection) to enhanced functional connectivity between brain regions involved in perception and encoding. It also suggests the involvement of a small brainstem structure, the locus coeruleus (LC), whose degeneration is increasingly associated with cognitive decline in aging. The process of encoding behaviorally relevant events into episodic memory thus involves large-scale, coordinated activation spanning cortical and subcortical regions.

Introduction

Attention can be allocated not only to spatial locations or stimulus features (Carrasco, 2011), but also to information presented at particular points in time (Rohenkohl et al., 2011; Nobre & van Ede, 2018). A growing literature shows that temporal selection enhances perceptual processing when events change in meaningful ways (e.g., Jefferies & Di Lollo, 2019) and when they require a response (Swallow & Jiang, 2010; Makovski et al., 2013; Yebra et al., 2019; Clewett et al., 2020). Detecting a behaviorally relevant item, for instance, can improve memory for concurrently presented but otherwise unrelated information (the attentional boost effect; Swallow & Jiang, 2010). This enhancement extends to multiple elements of an event across modalities and locations, even when those elements are task-irrelevant and ignored (Broitman & Swallow, 2020; Swallow & Jiang, 2014; Turker & Swallow, 2019). It is, however, limited in duration (Swallow & Jiang, 2011), binding only those items presented at the selected time into relational memories (Swallow & Atir, 2018; Turker & Swallow, 2019). These unique characteristics of temporal selection can be contrasted with those commonly observed in attention tasks, where competitive interactions within and between regions are prevalent (e.g., Johnson & Zatorre, 2005; Seidl et al., 2012; Silver et al., 2007).

Temporal selection may be supported by neuromodulatory mechanisms that increase gain in response to behaviorally relevant events or task boundaries (Jack et al., 2006; Lee et al., 2018; Swallow et al., 2012). Of particular interest is the locus coeruleus (LC), a brainstem structure whose activity briefly increases in response to changes in a task or in the environment (Clewett et al., 2020; Sara, 2009). Phasic LC responses correlate with pupil diameter (Murphy et al., 2014; Joshi et al., 2016) and are associated with target detection and orienting (Aston-Jones et al., 1994; Breton-Provencher & Sur, 2019). Its activity may also modulate declarative memory for behaviorally relevant moments (Swallow et al., 2019; Yebra et al., 2019). Because previous studies did not include baseline trials, it is unclear whether the reported effects reflect target-related facilitation or distractor-related inhibition. Furthermore, these studies utilized indirect measures of LC activity (pupil size) or a probabilistic atlas to identify the LC in participants,

making it difficult to pinpoint the source of the modulatory signals (cf. Turker et al., revising; Wang & Munoz, 2015).

MRI investigations of temporal selection have provided little insight into how it modulates neural processing. Prior work has shown that temporal selection broadly increases the BOLD signal in regions not directly involved in processing the target stimulus (e.g., auditory target detection boosts activity in V1; Jack et al., 2006; Swallow et al., 2012). However, differences in the magnitude of the hemodynamic response do not, on their own, entail differences in processing quality. To test whether temporal selection affects the depth and extent to which stimuli are processed, multivoxel pattern analysis is more suitable (Mahmoudi et al., 2012). It is also unclear to what extent temporal selection affects coordination among different brain regions. LC activation may lead to the dynamic reconfiguration of cortical functional networks (Shine et al., 2018; Li et al., 2019), which is thought to play role in the formation and maintenance of complex sensory representations (Fries, 2005, 2015). Enhanced functional connectivity (Friston, 2011) between the hippocampus and visual areas, in particular, has been linked to better working memory (Gazzaley et al., 2004) and episodic encoding (Ranganath et al., 2005).

We used fMRI to characterize the neural correlates of temporal selection, specifically examining responses of the visual cortex, hippocampus, and LC to images presented on their own or with auditory target or distractor tones. We tested whether target detection increased (1) phasic pupil responses and activity in anatomically defined LC; (2) the ability to classify patterns of BOLD activity associated with different categories of images; and (3) functional connectivity between visual regions and the hippocampus. We found evidence supporting each of these hypotheses.

Materials and Methods

Participants

21 right-handed individuals (15 female, 6 male, 19-40 years old, M = 21.48, SD = 4.86) participated in the study. They were screened for non-MRI compatible medical devices or body modifications, claustrophobia, movement disorders, pregnancy, mental illness, use of medication

affecting cognition, and color blindness. Consent was obtained at the beginning of the session and participants were debriefed at the end (all procedures approved by the Cornell University review board). Sample size was based on a previous study examining the effect of auditory target detection on visual cortical activity (Swallow et al., 2012), which reported effect sizes for a peak signal difference following target and distractor auditory tones of *Cohen's f* > 1.037. A sample size of 20 was selected to ensure that smaller effects between conditions and in other measures of connectivity and classification could be detected. With a sample of 20 and false positive rate of .05, a traditional one-way (3 levels) repeated measures analysis of variance has a power of .95 to detect an effect of *Cohen's f* > .378 (calculated using G*Power; Faul et al., 2007).

MRI and Pupillometry Data Acquisition

Magnetic resonance imaging was performed with a 3T GE Discovery MR750 MRI scanner (GE Healthcare, Milwaukee, WI) and a 32-channel head coil at the Cornell Magnetic Resonance Imaging Facility in Ithaca, NY. Participants laid supine on the scanner bed with their head supported and immobilized. Ear plugs and headphones were used to reduce scanner noise, allow the participant to communicate with the experimenters, and present auditory stimuli during the tasks. Visual stimuli were presented with a 32" Nordic Neuro Lab liquid crystal display (1920x1080 pixels, 60 Hz, 6.5 ms g to g) located at the head of the scanner bore and viewed through a mirror attached to the head coil.

Anatomical data were acquired with a T1-weighted MPRAGE sequence (TR = 7.7 ms; TE = 3.42 ms; 7° flip angle; 1.0 mm isotropic voxels, 176 slices). A second anatomical scan utilized a neuromelanin sensitive T1-weighted partial volume turbo spin echo (TSE) sequence (TR = 700 ms; TE = 13 ms; 120° flip angle; 0.430 x 0.430 mm in-plane voxels, 10 interleaved 3.0 mm thick axial slices; adapted from Keren et al., 2009). Slices for the TSE volume were oriented perpendicular to the long axis of the brain stem to provide high resolution data in the axial plane, where dimensions of the LC are smallest, and positioned to cover the most anterior portion of the pons. Multi-echo echo planar imaging (EPI) sequences were used to acquire functional data during the four task runs (TR = 2500 ms; TEs

= 12.3, 26.0, 40.0 ms; 83° flip angle; 3.0 mm isotropic voxels; 44 slices). Before the task scans, participants completed a single resting state scan with their eyes open and the lights on (612s; TR = 3.0 s; TEs=13, 30, 47 ms). Resting state data are reported elsewhere (Turker et al., revising) but were used to identify regions of interest for this study (see LC Functional Connectivity). Participants also completed additional scans at the end of the session that were not used in the study.

During the scans, pupil size and gaze location were acquired using an EyeLink 1000 Plus MRI Compatible eye tracker (SR-Research, Canada) for all but two participants. After the participant was positioned in the scanner, mirrors were adjusted to bring the right eye into view of the camera. Immediately prior to the resting state scan, thresholds defining pupil and corneal reflectance were automatically adjusted and a nine-point calibration routine was performed to determine the parameters needed to estimate gaze position. Calibration was validated and adjusted as necessary prior to each scan that included eye data measurement. On task scans, participants were instructed to fixate the central dot and minimize blinking.

MRI Data Preprocessing

All EPI data were denoised and processed using the standard ME-ICA pipeline, except as indicated (meica.py, Version 3.2, beta 1; Kundu et al., 2012; Kundu et al., 2013). First, the MPRAGE volume was skull stripped using FSL BET (b = 0.25). After matching the obliquity of the anatomical volume and EPI time series, motion was estimated from the first echo time series using 3dvolreg and the third volume as the target. Third, all EPI data were despiked and slice time acquisition differences were corrected using 3dTshift. Fourth, for each echo time series, the first two volumes were dropped and the remaining EPI data were registered to the third volume. Baseline intensity volume (s0), the t2* map volume (t2*), and the optimal combination volume time series (OCV) were then calculated. Fifth, registration and alignment transforms were applied to the EPI data and the pre-equilibrium volumes dropped in one step to align the data with the individual anatomical volume in its original acquisition space. Sixth, EPI data were denoised to identify and separate BOLD components from non-BOLD

components (Kundu et al., 2013). BOLD components were recombined to create the denoised data sets that were used in subsequent analyses. Finally, denoised EPI data were spatially aligned to the MNI_caez_2009 atlas for volume-wise group level analyses.

Region of Interest Identification

Individual MPRAGE scans were submitted to FreeSurfer's segmentation and surface-based reconstruction software (recon-all v5.3; http://surfer.nmr.mgh.harvard.edu/; Dale et al., 1999; Fischl et al., 1999) to label voxels according to each individual's anatomy. Labels for the fourth ventricle (4thV), hippocampus (HPC), motor cortex (MC), planum temporale for auditory cortex (AC), primary visual cortex (V1), V2, fusiform gyrus (FG), and parahippocampal gyrus (PG) were extracted and converted to volumetric ROIs using FreeSurfer and AFNI tools (Cox, 1996; Cox & Hyde, 1997; Gold et al., 1998). Separate ROIs were created for the left and right hemispheres. In addition, the HPC ROIs were divided into anterior and posterior portions at the anterior-posterior coordinate of their center of mass (aHPC, pHPC). Detailed methods for identifying the locus coeruleus (LC) are described in Turker et al. (revising). Briefly, bilateral LC ROIs were hand-drawn by two tracers on images acquired with neuromelanin sensitive TSE MRI sequences. Voxels included as LC by both raters were kept and then spatially aligned with the individual EPI and MPRAGE volumes after identification using a 12-parameter affine transformation.

Experimental Design and Statistical Analysis

Stimuli. 288 full color images of faces (48 female, 48 male), objects (48 cars, 48 chairs), and outdoor scenes (48 beaches, 48 forests) were acquired from personal collections and publicly available online databases (http://vision.stanford.edu/projects/sceneclassification/resources.html; Huang et al., 2007; Huang & Learned-Miller, 2014; Xiang et al., 2014). 24 images from each subcategory were used in the encoding task (eight in each tone type condition), each presented once per run (four repetitions in total). The rest of the images were used as foils in the recognition test. Scrambled images were generated from

these photographs by dividing them into 256 tiles and shuffling their location. The mean and variance of pixel intensity (luminance) was matched across images using the SHINE toolbox (Willenbockel et al., 2010).

Design and Procedure. In the four functional scans (407.5 s each) participants continuously performed simultaneous picture encoding and target detection tasks. On each 1.25 s long trial, one image (7 x 7 visual degrees; 256 x 256 pixels) was presented for 625 ms and immediately followed by another image for another 625 ms. On most trials, both images were scrambled (no-task trials, 164/run). On task trials (144/run), the first image was a photograph and the second was a scrambled version of that photograph. Participants attempted to memorize the photograph for a later memory test. The inter-trial interval was 0 ms, ensuring that there was an image on the screen throughout the task and that it changed every 625 ms (Figure 1A). A red fixation dot (0.25 visual degree diameter) appeared at the center of the screen throughout the task, including 7.5 s of pre-task fixation and 15 s of post-task fixation at the beginning and end of each run, respectively.

On some task trials a high (1200 Hz) or low (400 Hz) pitched auditory tone was played over the headphones at the same time a picture was presented (0 ms stimulus onset asynchrony). Sound levels were adjusted during the MPRAGE scan to ensure participants could hear both tones during scanning. Tone condition was held constant for each photograph. If the tone was a prespecified pitch (target) participants pressed a button with their right thumb. Participants made no overt response if the tone was not the pre-specified target pitch (distractor) or if no tone was presented on that trial. An equal number of task trials was assigned to each tone type (target, distractor, and no tone). Tones were never presented on no-task trials. Participants practiced the task with a different set of images before entering the scanner.

After scanning participants completed a two alternative forced choice recognition test on the pictures. On each trial two photographs were presented on the screen, one on the left side and one on the right. One of the pictures was presented during the encoding and detection task and the other was a new picture. Participants selected the ‘old’ photograph by pressing one of two keys on the keyboard.

Participants were then prompted to report their confidence by clicking on a line that appeared below the pictures. Participants were told to click on the far-left side of the line if they were guessing, the far-right side of the line if they were absolutely confident that they were correct, and at points in between to reflect degrees of intermediate levels of confidence. These were coded on a scale from 1 (lowest rating) to 100 (highest rating). A green + or a red - appeared next to indicate their accuracy. One participant did not complete the memory test due to a fire alarm, and a computer crashed during the memory test for one other participant.

This procedure resulted in a 6 x 3 design, with *picture type* (female face, male face, beach, forest, car, chair) and *tone type* (no-tone, distractor tone, target tone) as within-participants factors. There were 32 trials per picture-by-tone condition for a total of 576 task trials over four runs. Each run started with 7.5 s of fixation and ended with 15 seconds of fixation. Trial order and spacing were optimized using the AFNI function make_random_timing to produce four sequences that minimized the amount of unexplained variance in a simulated task. Task trials were separated by 0-12 non-task trials.

Behavioral Data Analysis. To examine the effect of tone type on memory we fit a binomial generalized linear mixed effects model (Bates et al., 2015) to the recognition accuracy data with tone type as a fixed effect and with random intercepts for participant, old image, and new image. A linear mixed effects model was fit to participants' recognition confidence ratings for correctly recognized images, with the same variables and random intercepts.

Tonic and Phasic Pupil Size Estimation and Analysis. Before estimating tonic and phasic pupil dilation on a trial-by-trial basis, pupil data for each participant and task run were preprocessed using the EyeLink DataViewer application (SR-Research, Canada), the FIRDeconvolution toolbox (Knapen et al., 2016), and custom routines. In brief, linear interpolation was used to estimate pupil size during blinks flagged by the EyeLink software and extended to include 100 ms margins before and after the blink. High-pass (0.1 Hz) and low-pass (10 Hz) Butterworth filters were applied, after which the data were downsampled from

1000 Hz to 100 Hz. Noise associated with the end of blinks and saccades was then removed as follows. For each participant, mean pupil diameter was calculated for every sample during the 6 s time windows following a blink and following a saccade (*pupil response*). A double gamma impulse response function (IRF) was then fit to the pupil response following a blink. A single gamma IRF was fit to the pupil response following a saccade. The blink and saccade IRFs were convolved with blink and saccade ends to create individually tailored nuisance regressors, and a cleaned data set was then acquired by using the residuals from a linear model describing measured pupil responses as a function of these nuisance regressors. Finally, the previously filtered out slow drift was added back into the data, as this is a meaningful characteristic of pupil size over time.

Tonic pupil size and phasic pupil response were then estimated for each participant and trial of the encoding and detection task. For every trial, tonic pupil size was defined as the mean pupil size in the 500 ms window preceding the trial. In addition, uncorrected phasic pupil responses were defined as the area under the curve (AUC; using Simpson's rule) defined by the double gamma IRF that best fit the pupil time series during the 2 s interval following trial onset. These values were then z-scored by subtracting the individual participant's mean and dividing by the standard deviation to produce the *scaled phasic pupil response*. A linear mixed effects model with random intercepts for participant indicated a significant, negative relationship between the scaled phasic pupil responses and tonic pupil size, $\beta = -.282$, 95% CI = [-.320, -.244], $t(2446) = -14.53$, $p < .001$. Residuals from this model were used in subsequent analyses and are referred to as *phasic pupil responses* for simplicity.

The effects of tone type on phasic pupil responses during encoding were evaluated by fitting a linear mixed effects model to the phasic pupil responses (averaged over presentations of an image) with tone type as a fixed effect and random intercepts for the image. Random intercepts for participant had near-zero variance and were thus removed from the final model.

Univariate Analysis. Following pre-processing, EPI volumes with motion greater than 0.3 mm were excluded and the data were spatially smoothed using a Gaussian kernel until blur reached a full-width-

half-maximum of 5.0 mm (3dBlurtoFWHM). To better estimate activity in the LC, voxels in the neighboring fourth ventricle, labeled with FreeSurfer (recon-all), were excluded from smoothing and subsequent analyses. In addition, masks defining the spatial extent of the brain in the aligned anatomical and EPI data sets, excluding the fourth ventricle, were applied to the EPI data. Data were scaled to a mean of 100 and a range of 0-200 to allow interpretation of beta weights as percent change.

Responses to events of different types were estimated for each voxel using 3dDeconvolve. All models included six motion regressors and 3rd order polynomial drift in baseline as nuisance variables. Regressors of interest were created by convolving a delta function for each event of interest with the two-parameter SPMG2 hemodynamic response function (HRF; Henson et al., 2002). In the univariate encoding and detection task analyses, regressors were included for each combination of tone type and picture type, for a total of 18 regressors of interest. When using the SPMG2 HRF, 3dDeconvolve produces two beta estimates for each condition. These were used to estimate the first 5 timepoints (12.5 s) of the hemodynamic response to each of the 18 conditions for subsequent group level analyses.

Univariate analyses of the ROIs were performed by extracting the mean estimated HRF across voxels located within the boundaries of the ROIs for each of the 18 conditions. For each ROI, estimated HRFs were additionally averaged across picture type and analyzed in R with a linear mixed effects model that included tone type, time (timepoints 0 s – 12.5 s), hemisphere (left, right), and all interactions as fixed effects and random intercepts for participant and picture type. The one exception was the LC ROI, which was collapsed across hemispheres. Models were simplified by excluding the interactions with hemisphere for ROIs that did not show a hemisphere by tone type interaction (all but MC). To characterize the effects of tone type over time in each ROI, general linear tests comparing activity across encoding conditions were then performed for each time point.

Whole brain, group-level univariate analyses were performed to characterize the effects of target and distractor tones on activity throughout the brain. Voxels for which there was a significant interaction between tone type and time were identified in an analysis of variance, with tone type, picture type, and time (timepoints 0 s – 12.5 s) as factors, using 3dMVM (Chen et al., 2015). To further characterize the

interaction between tone type and time, the statistical map for this interaction was thresholded at $q < .001$ to create a mask of voxels whose hemodynamic response significantly differed across the three tone type conditions. Post hoc paired t-tests (3dttest++) on voxels within the tone type by time interaction mask were performed on timepoints 2.5 s – 7.5 s of the estimated HRF to target vs. distractor tones, target tones vs. no tones, and distractor tones vs. no tones. Statistical and cluster size thresholds were used to correct for multiple comparisons based on simulations that used spatial auto-correlation functions (using AFNI function 3dClustSim; Cox et al., 2017).

Trial-Specific Activity Estimation. Trial-specific activity was estimated by fitting a separate linear model for each trial of the encoding and detection task using the least square-separate approach (LSS; Mumford et al., 2012). The deconvolution was performed using AFNI's 3dDeconvolve with the SPMG2 option, such that each single-trial response was modeled by two regressors (a gamma response function and its time derivative). Similarly, the combined responses on all other trials were modeled by two nuisance regressors. In addition, the design matrix included the same motion and drift nuisance variables used in the univariate model described above. The single-trial gamma function estimates were then saved, resulting in 576 (4 runs x 144 trials) beta maps.

Beta Series Correlation Analysis. The effects of tone type on functional connectivity between the HPC and visual ROIs was estimated using beta series correlations (Rissman et al., 2004; Cisler et al., 2014; Geib et al., 2017) generated from the trial-specific activity estimates. To avoid introducing distortions, we did not subtract the mean pattern from each voxel or scale the data prior to computing these values (Garrido et al., 2013).

First, we concatenated the beta weights of trials sharing the same tone type condition (resulting in three series per voxel with 192 elements each). To obtain ROI-to-ROI functional connectivity estimates, separately for each tone type condition, we generated a mean beta series (obtained by averaging the series across voxels) for each ROI and computed pairwise Fisher-transformed Pearson correlations between

those. First, we fit a linear mixed effect model to the correlation coefficients, with tone type and ROI pair as fixed factors and random intercepts for participant. We then fit a separate linear model for each ROI pair that included a visual and a hippocampal ROI, with tone type as a fixed effect and random intercepts for participant to test our specific hypotheses regarding visuo-hippocampal functional connectivity. Holm-corrected general linear tests comparing the different levels of tone type were performed for each model.

ROI-to-voxel beta correlation analyses were also performed for each hippocampal ROI to gain a finer grained understanding of the effects of tone type on connectivity. Fisher-transformed Pearson correlations between the mean beta series of the seed ROI and each voxel in all visual and hippocampal ROIs (including those in the seed ROI) were then calculated for each tone type and participant. Linear mixed effects models (3dLME) with tone type as a fixed effect and random intercepts for participant were then fit to the ROI-to-voxel correlations. General linear tests contrasted the target with the distractor and no tone conditions at each voxel. Candidate clusters in the resulting maps were identified after accounting for spatial autocorrelation in the data (estimated with 3dFWHMx) and by thresholding based on a minimal cluster size and maximal p-value (voxel edges must touch, $a = .05$, uncorrected $p = .05$; 3dClustSim and 3dClusterize; Cox et al., 2017). This was followed by confirmatory analyses in which each cluster was treated as an ROI. The correlation between the average beta series of the ROI and that of the respective hippocampal seed was computed separately for each participant. A linear mixed effects model, with tone type as a fixed effect and random intercepts for participant, was fit to the correlations. Follow-up general linear tests contrasting the tone type conditions were then performed. This was done to ensure that the clustering procedure did not produce spurious clusters.

Support Vector Classification. Linear support vector machine (SVM) classification (Suykens & Vandewalle, 1999; Hsu & Lin, 2002) probed the effects of target tone detection on picture category decoding accuracy in the visual and hippocampal ROIs (V1, V2, FG, PG, aHPC, and pHPC).

Each trial-wise beta map was assigned one of six labels indicating the type of picture presented on that trial. Individual beta series maps were standardized across trials such that each voxel had a mean of zero and a standard deviation of one. Classification accuracy was estimated for each tone type and ROI using repeated 4-fold cross validation. Balanced training sets of 144 trial-wise beta maps were randomly drawn 30 times (the remaining 48 trials in each iteration were reserved as a test set); on each iteration a new linear SVM (C=1, one-vs-one multiclass, default implementation in scikit-learn; Pedregosa et al., 2011) was fit to the training set and applied to the test set to obtain a confusion matrix and a classification accuracy estimate for each tone type condition. These were averaged across the 30 iterations to produce one estimate and one confusion matrix for each combination of participant, tone type, and ROI. The effects of tone type on mean classification accuracy were evaluated for each ROI using linear mixed effects models with tone type as a fixed effect and random intercepts for participant.

LC Functional Connectivity. ROIs that exhibited functional connectivity with LC during rest were identified. As described elsewhere (Turker et al., revising), intrinsic functional connectivity (iFC) maps were created for each participant using denoised multi-echo data and the participant's individually defined LC ROI. A group-level iFC map was created using voxel-wise t-tests (3dttest++) and a one-sided clustering procedure at $p = .01$ and FDR = .018 (3dClusterize; corrected for multiple comparisons using the false discovery rate (FDR = .02; Geonvese et al., 2002). Twenty ranked peaks were extracted from the group iFC map (3dmaxima) with a minimal distance of 18 mm (6 voxels) between peaks. Next, 6 mm spherical ROIs were constructed around those peaks and a final set of twenty ROIs was obtained by intersecting the spheres with the group-level iFC map thresholded at $p < .001$, $q < .004$, producing the final LC-iFC ROIs.

Statistical Software

All group-level analyses were performed in R v3.6.1 (R Core Team, 2013) or in AFNI v16.2.07 (using 3dLME or 3dMVM). Unless otherwise noted, linear mixed effects models were fit using lme4 v1.1.21

(Bates et al., 2015). Type III (Satterthwaite's method) ANOVA tables were obtained using the 'joint_tests' function in the package emmeans v1.3.5.1 (Length, 2019). General linear tests were performed and uncorrected confidence intervals were obtained using the emmeans functions 'contrasts' and 'confint.' In all analyses, Holm-Bonferroni adjusted p-values were computed separately for each set of tone type comparisons. Confidence intervals, where reported, are uncorrected.

Results

Behavioral Task Performance and Whole Brain Analysis

Participants accurately performed the detection task, pressing the button for 97.5% of the targets, 5.1% of the distractors, and 0.2% of the no tone trials. Incorrect button presses were more likely to follow a distractor tone than no tone, $t(20) = -4.16$, $p < .001$, $d = 1.25$.

The tone type condition did not significantly influence memory for the images, $F(2, \text{inf}) = 1.08$, $p = .34$. However, it did influence the confidence with which images were correctly recognized, $F(2, 1660.8) = 6.81$, $p = .001$. Participants reported higher levels of confidence for correctly recognized images paired with a target than for those paired with a distractor ($M_{Diff} = 5.58$), 95% CI = [1.83, 9.32], $t(1899.08) = 3.57$, $p = .001$, $d = .57$, or presented without a tone ($M_{Diff} = 4.10$), 95% CI = [0.23, 7.97], $t(1573.68) = 2.54$, $p = .023$, $d = .44$, but confidence did not differ between correctly recognized images presented in the distractor and no tone conditions ($M_{Diff} = 1.48$), 95% CI = [-2.38, 5.34], $t(1578.25) = 0.92$, $p = .359$, $d = .10$. Target detection during image encoding thus increased the confidence with which those images were later correctly recognized, and distractor rejection did not significantly interfere with the encoding of a concurrently presented image (Figure 1B).

Whole brain analyses revealed that auditory target detection influenced BOLD activity in regions spanning medial occipital, medial parietal, anterior cingulate, superior temporal, middle frontal, and subcortical areas, including thalamus (time X tone type interaction, $F > 3.483$, $q < .001$). In many of these regions, activity was initially higher on the target tone trials than in the other two types of trials, though this relationship reversed at subsequent timepoints (Figure 1C, top and middle row). Relative to no tone

trials, distractor trials elicited a response that was smaller in magnitude in most regions (Figure 1C, bottom row).

Locus Coeruleus and Phasic Pupil Responses

To test whether the LC system could be involved in temporal selection, we examined whether target tone trials evoked greater phasic pupil responses and LC signal changes than did distractor tone and no tone trials. Consistent with this possibility, BOLD responses in the hand-traced LC ROIs exhibited an interaction between tone type and time, $F(10, 2225) = 8.28$, $p < .001$, reflecting greater increases in activity at 2.5 s on target tone trials than on distractor trials, $t(2225) = 6.51$, $p < .001$, $d = .48$, and no tone trials, $t(2225) = 7.03$, $p < .001$, $d = .58$ (Figure 2). Phasic pupil responses also varied across tone type conditions, $F(2, 2124.6) = 336.98$, $p < .001$: they were more positive on trials that included a target tone than on trials that included a distractor tone ($M_{Diff} = 0.47$), 95% CI = [0.37, 0.56], $t(2110) = 11.45$, $p < .001$, $d = .25$, or no tone ($M_{Diff} = 1.05$), 95% CI = [0.96, 1.15], $t(2102) = 25.86$, $p < .001$, $d = .39$. They were also greater on distractor tone trials than on no tone trials ($M_{Diff} = 0.59$), 95% CI = [0.49, 0.68], $t(2160) = 14.44$, $p < .001$, $d = .13$ (Figure 2). These data demonstrate increased activity of the LC system on trials that require a response and provide no evidence for inhibitory effects of distractor rejection on this system.

BOLD Responses in Perceptual and Motor Regions

To examine the effects of temporal selection on the processing of episodic information, planned analyses tested whether target detection modulated BOLD magnitude regions involved in stimulus processing and encoding and response generation: bilateral MC, V1, V2, FG, PG, aHPC, and pHPC. Except for MC, there were no interactions between hemisphere and tone type, largest $F(2, 4475) = 1.80$, $p = .165$, or between tone type, time, and hemisphere, largest $F(10, 4475) = 1.77$, $p = .061$. In MC, at 2.5 s, BOLD activity was greater on target trials than on distractor and no tone trials in the left hemisphere, smallest $t(2225) = 13.55$, $p < .001$, $d = .98$.

The interaction between tone type and time was significant in V1, V2, FG, PG, aHPC, and pHPC, smallest $F(10, 4492) = 12.48$, $p = .001$ for FG. Consistent with earlier reports (Swallow, Makovski & Jiang, 2012), V1 showed a larger initial increase in activity on target trials than on distractor and no tone trials, smallest $z = 7.81$, $p < .001$, $d = .43$. Similar increases were also observed in V2, FG, and PG, smallest $z = 2.60$, $p = .019$, $d = .15$. Additionally, in all cases BOLD activity showed a steeper drop-off in magnitude on target trials relative to distractor and no tone trials (Figure 3).

The HPC generally showed larger decreases in activity on target trials than on distractor and no tone trials. This was true of both the anterior and posterior HPC, which at 5 s were more strongly deactivated on target trials than on distractor trials, smallest $z = 4.50$, $p = .001$, $d = .28$, and no tone trials, smallest $z = 11.81$, $p < .001$, $d = .751$. However, activity in the aHPC decreased more rapidly than it did in the pHPC: at 2.5 s, activity in the aHPC was lower on target than on distractor, $z = 3.12$, $p = .002$, $d = .18$, and no tone, $z = 8.54$, $p < .001$, $d = .52$, trials, whereas activity in pHPC was higher on target trials than on distractor trials, $z = 3.49$, $p = .001$, $d = .20$ (Figure 3). Target detection therefore modulated the magnitude of activity in regions involved in representing visual and episodic information.

Decodable Stimulus Information in the Visual Cortex and Hippocampus

Increased BOLD activity on target trials does not necessarily imply that processing in these regions is enhanced. Therefore, we conducted analyses of picture category classification accuracy using voxel-wise patterns of activity in the visual and hippocampal ROIs. These revealed a main effect of tone type in V2, FG, PG, and pHPC, smallest $F(2, 92) = 4.59$, $p = .013$ for pHPC, but not in V1 or aHPC, largest $F(2, 92) = 1.36$, $p = .261$ for V1. Follow up analyses indicated that classification accuracy was higher on target trials than on no tone trials in V2, FG, PG, and pHPC, smallest $t(92) = 3.01$, $p = .01$, $d = .96$ for pHPC. Classification accuracy was also higher on target compared to distractor tone trials in FG $t(92) = 2.60$, $p = .011$, $d = .62$, PG, $t(92) = 2.09$, $p = .039$, $d = .52$, and marginally in V2, $t(92) = 2.17$, $p = .065$, $d = .44$. Similarly, relative to no tone trials, distractor trials significantly enhanced classification accuracy in FG and PG, smallest $t(92) = 2.764$, $p = .014$, $d = .87$ for the latter, and showed a marginal

effect in V2, $t(92) = 1.88$, $p = .065$, $d = .43$ (Figure 3). Thus, many of the regions whose activity was enhanced by target detection also exhibited within-ROI patterns of activity that better correlated with picture category on these trials than on distractor or no tone trials. This was particularly true in higher level visual areas, FG and PG, which should be more tuned to picture categories than V1 and V2.

Visuo-Hippocampal Functional Connectivity

To examine the effect of tone type on inter-areal communication, functional connectivity (FC) between all visual and hippocampal ROIs was quantified with beta series correlations (Figure 4). Overall, tone type and ROI pair influenced FC, $F(2) = 46.65$, $p < .001$ and $F(77) = 122.24$, $p < .001$, but did not interact, $F(154) = 0.33$, $p > .999$. General linear tests indicated that FC was enhanced on target trials relative to both distractor trials, $z = 9.48$, $p < .001$, $d = .50$, and no tone trials, $z = 6.36$, $p < .001$, $d = .32$. FC also was higher on no tone trials than on distractor tone trials, $z = 3.12$, $p = .002$, $d = .24$. Thus, relative to the no tone trials, target tones increased correlations between ROIs while distractor tones decreased them.

In follow-up analyses a separate model was fit to each visuo-hippocampal pair (i.e., the lower left quadrants of the connectivity matrices in Figure 4) and a series of planned comparisons contrasted the tone type conditions for each of those. Tone type modulated FC between l-pHPC and l-FG, $F(2, 36) = 4.48$, $p = .018$, l-pHPC and l-V2, $F(2, 36) = 3.75$, $p = .033$, and r-aHPC and l-FG, $F(2, 36) = 4.54$, $p = .018$. For these three ROI pairs, FC was stronger on target trials than on distractor trials, smallest $t(36) = 2.68$, $p = .033$, $d = .52$ for the l-pHPC and l-V2 pair. FC was also stronger on target trials than no tone trials for both the l-pHPC and l-FG and the r-aHPC and l-FG pairs, smallest $t(36) = 2.51$, $p = .038$, $d = .49$ (Figure 4). There were no significant differences in visuo-hippocampal FC on distractor and no tone trials, largest $t(36) = 1.41$, $p = .334$, $d = .338$ for the l-aHPC and r-pHPC pair. Tone type did not significantly affect FC in the remaining visuo-hippocampal ROI pairs, largest $F(2, 36) = 2.98$, $p = .063$ for r-aHPC and r-FG.

To more precisely identify the regions whose functional connectivity with the HPC changed with tone type, we calculated ROI-to-voxel FC maps and extracted candidate clusters by contrasting the tone type conditions. We refer to these clusters by the ROI seed that generated them and the anatomical area that they overlapped with most (e.g., l-pHPC<->r-V2 refers to a cluster largely overlapping with r-V2 that was correlated with l-pHPC). No clusters were found when the distractor condition was contrasted with the no tone baseline, indicating that the distractor tones did not reliably alter FC between the hippocampus and visual cortex. However, three clusters (l-pHPC<->l-FG, l-pHPC<->r-FG, and r-aHPC<->l-FG) were identified when target trials were contrasted with no tone trials and seven clusters were identified when target trials were contrasted with distractor trials (Table 1 and Figure 5). Both sets of clusters spanned voxels throughout visual cortex.

Confirmatory analyses on clusters identified in the target versus distractor contrast were performed by averaging FC across all voxels within a cluster and then testing the effect of tone type in a linear mixed effects model. This analysis indicated that all pairs showed an effect of tone type on FC, smallest $F(2, 36) = 3.77$, $p = .033$ for l-aHPC<->r-FG, except l-aHPC<->l-FG, $F(2, 36) = 3.02$, $p = .061$. In those pairs showing an effect of tone type, FC was higher on target trials than on distractor trials, smallest $t(36) = 2.61$, $p = .039$, $d = .54$ for l-aHPC<->r-FG. FC was also higher on target trials than on no tone trials in the pairs l-pHPC<->r-V2, l-pHPC<->l-V2, and r-aHPC<->l- FG, smallest $t(36) = 2.39$, $p = .045$, $d = .46$ for l-pHPC<->l-V2. No differences in FC were found between the distractor and no tone trials in any of the clusters, largest $t(36) = 1.12$, $p = .273$, $d = .29$ (Figure 5).

To test the hypothesis that increased visuo-hippocampal coordination during target tone trials is associated with better visual processing, we treated the ROI-to-voxel FC clusters as ROIs in a picture category classification analysis. Tone type affected classification accuracy in all clusters except r-pHPC<->l-V2, smallest $F(2, 36) = 3.53$, $p = .04$. This effect reflected greater accuracy on target trials than on no tone trials, smallest $t(36) = 2.65$, $p = .035$, $d = .86$. Only the l-pHPC<->r-V2 cluster showed higher classification accuracy on target trials than on distractor trials, $t(36) = 2.54$, $p = .031$, $d = .74$. Accuracy

was higher on distractor trials than on no tone trials in r-pHPC<->r-FG and r-aHPC<->l-FG, smallest $t(36) = 2.48$, $p = .036$, $d = .77$ for the former (Figure 5).

Functional Connectivity of Regions Associated with LC During Rest

If the LC influences activity in regions involved in episodic encoding, then FC between HPC and regions whose activity is modulated by the LC during rest may be greater on target trials (when LC activity is strongest) than on distractor and no tone trials. To test this hypothesis, a set of 20 regions whose activity was associated with LC activity during a separate resting state scan—referred to as LC-iFC ROIs—were identified (see methods and Turker et al., revising; Table 2) and their FC to the four hippocampal seeds was examined. A separate model was fit for each of l-pHPC, r-pHPC, l-aHPC, and r-aHPC. Main effects were found for tone type, smallest $F(2, 1062) = 10.11$, $p < .001$ for l-aHPC, and region, smallest $F(19, 1062) = 13.05$, $p < .001$ for r-pHPC, but the two did not interact, largest $F(38, 1062) = 0.49$, $p = .996$ for l-pHPC. General linear tests that collapsed across LC-iFC ROIs showed higher FC on target trials than on distractor trials, smallest $t(1062) = 4.45$, $p < .001$, $d = .41$ for l-aHPC, and no tone trials, smallest $t(1062) = 2.45$, $p = .029$, $d = .24$ for l-pHPC. FC to l-pHPC was lower on distractor trials than on no tone trials, $t(1062) = 2.06$, $p = .039$, $d = .32$. However, tone type did not significantly influence FC between the LC and the LC-iFC ROIs, $F(2, 1062) = .093$, $p = .911$.

Discussion

In this fMRI study we investigated the mechanisms underlying temporal selection using a target detection task. We examined how visual regions, the hippocampus, and the LC respond to pictures presented concurrently with an auditory target tone (requiring a motor response), a distractor tone (requiring no response), or no tone—a variant of the attentional boost paradigm (Swallow & Jiang, 2010, 2014). The inclusion of a no tone baseline condition, which was absent in previous fMRI studies of similar effects, allowed us to test for both target-induced enhancement and distractor-induced disruption of encoding, perceptual processing, and functional connectivity. We found that, relative to the no-tone

and distractor conditions, target tones enhanced picture recognition confidence, the magnitude of phasic pupil responses, activity in LC and in visual regions, visuo-hippocampal functional connectivity, and picture category classification accuracy from multivoxel patterns in FG and PG.

Consistent with existing evidence suggesting that target detection enhances episodic encoding (Swallow & Jiang, 2013), we found that recognition confidence ratings for correctly remembered images was greater when those were paired with a target tone as opposed to a distractor tone. Indeed, previous work on the attentional boost effect has shown that target detection can enhance both recollection and familiarity of images presented during the encoding task—particularly when they are repeated multiple times or presented for sufficiently long periods of time (Broitman & Swallow, 2020). The absence of significant differences in recognition performance is consistent with an earlier MRI study (Swallow et al., 2012) and may be a consequence of the relatively long inter-trial intervals that had to be used in this study (cf. Mulligan & Spataro, 2015).

Throughout the visual areas examined, activity in the target tone condition was higher relative to the distractor and baseline conditions. This pattern was also present in other regions involved in orienting to relevant stimuli. Those included the thalamus, whose higher-order nuclei play a critical role in attention and in the maintenance of sensory representations in awareness and working memory (Saalmann & Kastner, 2011), and the LC, whose noradrenergic projections regulate gain throughout the cortex (Aston-Jones & Cohen, 2005; Servan-Schreiber et al., 1990) and which tends to exhibit larger responses following salient events (Bouret & Sara, 2005). Interestingly, the opposite pattern was observed in aHPC and pHPC: activity decreased on target tone trials relative to the two other conditions. Because the images were presented four times, these decreases may reflect greater repetition suppression of hippocampal activity for target-paired images than for other images (Henson & Rugg, 2003; Summerfield et al., 2008; Larsson & Smith, 2011; Kim et al., 2020). Measures of repetition suppression tend to positively correlate with subsequent recognition memory (e.g., Pihlajamaki et al., 2010) and with functional connectivity between the hippocampus, visual cortex, and prefrontal cortex (Zweynert et al., 2011). Though our

results, as a whole, hint at a relationship between hippocampal repetition suppression and the effects of temporal selection on recollection, additional work is needed to confirm this possibility.

Changes in inter-areal coordination are thought to play a role in the maintenance and encoding of neural representations (Singer, 2013; Fries, 2005, 2015; Bonnefond et al., 2017; Moyal & Edelman, 2019; Moyal et al., 2020). Enhanced cortico-hippocampal functional connectivity, in particular, has been associated with working memory maintenance and long-term memory encoding (e.g., Gazzaley et al., 2004; for a review, see Poch & Campo, 2012). Temporal selection could influence perception and memory in a similar fashion, by enhancing communication between perceptual and medial temporal regions in critical moments (e.g., in response to target detection). Our findings are compatible with this idea: functional connectivity between the HPC and the visual cortex was higher on target tone trials relative to both distractor and no tone trials. This effect was most pronounced for the left pHPC, which showed widespread, bilateral increases in functional connectivity with clusters in the ventral visual cortex—V1, V2, and FG. The aHPC exhibited a similar correlation with FG, but less so with V1 and V2, potentially reflecting the differential connectivity of the anterior and posterior HPC in humans (Poppenk et al., 2013).

The target-related enhancement of visuo-hippocampal connectivity was accompanied by a boost in picture category classification accuracy throughout the ventral visual stream and in pHPC relative to the baseline condition. A similar enhancement was found relative to the distractor tone condition in FG and PG. This suggests that the effect of temporal selection on subsequent recognition confidence could reflect improved perceptual encoding in higher level visual areas. Surprisingly, a smaller increase in classification accuracy was also observed on distractor tone trials relative to the no-tone baseline in FG and PG. Thus, although exposure to any auditory tone in this task may enhance the quality of visual information processing, target detection provides an additional boost. A similar pattern was also found in smaller ventral visual clusters that exhibited greater functional connectivity with the hippocampus on target compared to distractor tone trials. This suggests a possible relationship between these two effects.

Future studies may address the question of whether shifts in long-range coordination are directly related to the quality and extent of perceptual processing and memory encoding (Moyal & Edelman, 2019).

Phasic LC responses have been hypothesized to facilitate the updating of representations and contribute to temporal selection (Swallow & Jiang, 2010; Swallow et al., 2012) by enhancing perceptual processing following target detection independent of modality or spatial location (Swallow & Jiang, 2013; Swallow & Jiang, 2014; see also Nieuwenhuis et al., 2005; Bouret & Sara, 2005). Our results are consistent with this view. They suggest a role for the LC in mediating the effects of temporal selection on perception and memory via the strengthening or reorganization of functional networks. In this study, both LC activity and phasic pupil responses (which may correlate; Joshi et al., 2016) increased on target tone trials relative to both distractor and no tone trials. These differences were observed despite the fact that each condition was equally likely. However, whereas LC activity did not increase on distractor trials relative to no tone trials, pupil diameter did. There therefore may be other cognitive and neural factors that contribute to pupil size in addition to LC activity (e.g., cognitive effort; Van der Wel & Van Steenbergen, 2018). Additionally, regions that were highly correlated with the LC during rest were also more correlated with HPC on target trials than on distractor trials during the encoding and detection task. This hints at a possible link between target-related LC responses and enhanced hippocampal functional connectivity, which can be addressed directly in future studies.

In the encoding and detection task, target detection differs from distractor rejection in both temporal selection and in the motor response demands. However, the effects of target detection on activity in visual cortex, pupil responses, and memory can occur in the absence of an overt motor response and are absent when motor responses are self-generated (Jack et al., 2006; Swallow & Jiang, 2012; Swallow et al., 2012; Makovski et al., 2013; Swallow et al., 2019), making it unlikely that the results we report in visual processing can be attributed to motor responses. Similarly, the behavioral inhibition in the distractor tone condition did not lower BOLD magnitude or classification accuracy relative to the baseline, ruling out the possibility of a disruptive effect of response inhibition on

processing (as in inhibition-induced forgetting, which has been tied to fluctuations in ventrolateral PFC activity, which were not found here; Chiu & Egner, 2015) and confirming that our findings truly reflect a target-induced boost (Swallow & Jiang, 2014).

Taken together, our data suggest that temporal selection facilitates the processing and encoding of information presented at the same time as stimuli that require a response. These events enhance perceptual processing and functional connectivity between the hippocampus and the ventral visual cortex. Both effects could be related to the stronger LC responses observed following auditory target detection. Our results and interpretation are compatible with the emerging view that, by increasing gain throughout the thalamocortical network at opportune moments, the phasic release of norepinephrine from the LC may facilitate functional network reorganization and promote more integrated, information-rich dynamics. Theoretical models and empirical findings have linked higher gain to increases in the topological complexity and variability of population activity (Shine et al., 2018; Moyal & Edelman, 2019) as well as to enhanced inter-regional information transfer (Li et al., 2019). Future work may combine functional neuroimaging, electrophysiology, and computer simulations to explore these possibilities and provide a precise account of the mechanisms underlying temporal selection. This work can further clarify the effects of attending to behaviorally relevant moments on neural dynamics, information representation, and incidental encoding.

References

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, *28*, 403-450.

Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994). Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *Journal of Neuroscience*, *14*(7), 4467-4480.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, Articles, 67*(1), 1–48

Bonnefond, M., Kastner, S., & Jensen, O. (2017). Communication between brain areas based on nested oscillations. *eneuro*, *4*(2).

Bouret, S., & Sara, S. J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends in neurosciences*, *28*(11), 574-582.

Breton-Provencher, V., & Sur, M. (2019). Active control of arousal by a locus coeruleus GABAergic circuit. *Nature neuroscience*, *22*(2), 218-228.

Broitman, A. W., & Swallow, K. M. (2020). The effects of encoding instruction and opportunity on the recollection of behaviourally relevant events. *Quarterly Journal of Experimental Psychology*, *73*(5), 711-725.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision research*, *51*(13), 1484-1525.

Chen, G., Saad, Z. S., Adleman, N. E., Leibenluft, E., & Cox, R. W. (2015). Detecting the subtle shape differences in hemodynamic responses at the group level. *Frontiers in Neuroscience*, *9*, 375.

Chiu, Y. C., & Egner, T. (2015). Inhibition-induced forgetting results from resource competition between response inhibition and memory encoding processes. *Journal of Neuroscience*, *35*(34), 11936-11945.

Cisler, J. M., Bush, K., & Steele, J. S. (2014). A comparison of statistical methods for detecting context-modulated functional connectivity in fMRI. *Neuroimage*, *84*, 1042-1052.

Clewett, D., Gasser, C., & Davachi, L. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature Communications*, *11*(1), 1-14.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, *29*(3), 162-173.

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). FMRI clustering in AFNI: false-positive rates redux. *Brain connectivity*, *7*(3), 152-171.

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, *10*(4-5), 171-178.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179-194.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage*, *9*(2), 195-207.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, *9*(10), 474-480.

Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron*, *88*(1), 220-235.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity*, *1*(1), 13-36.

Garrido, L., Vaziri-Pashkam, M., Nakayama, K., & Wilmer, J. (2013). The consequences of subtracting the mean pattern in fMRI multivariate correlation analyses. *Frontiers in neuroscience*, *7*, 174.

Gazzaley, A., Rissman, J., & D'esposito, M. (2004). Functional connectivity during working memory maintenance. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(4), 580-599.

Geib, B. R., Stanley, M. L., Dennis, N. A., Woldorff, M. G., & Cabeza, R. (2017). From hippocampus to whole-brain: The role of integrative processing in episodic memory retrieval. *Human brain mapping*, *38*(4), 2242-2259.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, *15*(4), 870-878.

Gold, S., Christian, B., Arndt, S., Zeien, G., Cizadlo, T., Johnson, D. L., ... & Andreasen, N. C. (1998). Functional MRI statistical software packages: a comparative analysis. *Human brain mapping*, *6*(2), 73-84.

Henson, R. N., Price, C. J., Rugg, M. D., Turner, R., & Friston, K. J. (2002). Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *Neuroimage*, *15*(1), 83-97.

Henson, R. N. A., & Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, *41*(3), 263-270.

Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, *13*(2), 415-425.

Huang, G. B., & Learned-Miller, E. (2014). Labeled faces in the wild: Updates and new reporting procedures. *Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep*, 14-003.

Huang, G.B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007, October). Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. *University of Massachusetts, Amherst, Technical Report* 07-49.

Jack, A. I., Shulman, G. L., Snyder, A. Z., McAvoy, M., & Corbetta, M. (2006). Separate modulations of human V1 associated with spatial attention and task structure. *Neuron*, *51*(1), 135-147.

Jefferies, L. N., & Di Lollo, V. (2019). Sudden events change old visual objects into new ones: A possible role for phasic activation of Locus Coeruleus. *Psychological science*, *30*(1), 55-64.

Johnson, J. A., & Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. *Cerebral Cortex*, *15*(10), 1609-1620.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*(1), 221-234.

Keren, N. I., Lozar, C. T., Harris, K. C., Morgan, P. S., & Eckert, M. A. (2009). In vivo mapping of the human locus coeruleus. *Neuroimage*, *47*(4), 1261-1267.

Kim, K., Hsieh, L. T., Parvizi, J., & Ranganath, C. (2020). Neural Repetition Suppression Effects in the Human Hippocampus. *Neurobiology of Learning and Memory*, 107269.

Knapen, T., de Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PloS one*, *11*(5), e0155574.

Kundu, P., Inati, S. J., Evans, J. W., Luh, W. M., & Bandettini, P. A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *Neuroimage*, *60*(3), 1759-1770.

Kundu, P., Brenowitz, N. D., Voon, V., Worbe, Y., Vértes, P. E., Inati, S. J., ... & Bullmore, E. T. (2013). Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proceedings of the National Academy of Sciences*, *110*(40), 16187-16192.

Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cerebral Cortex*, *22*(3), 567-576.

Lee, T. H., Greening, S. G., Ueno, T., Clewett, D., Ponzio, A., Sakaki, M., & Mather, M. (2018). Arousal increases neural gain via the locus coeruleus–noradrenaline system in younger adults but not in older adults. *Nature Human Behaviour*, *2*(5), 356-366.

Length, R. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package v ersion 1.4.4. https://CRAN.R-project.org/package=emmeans

Li, M., Han, Y., Aburn, M. J., Breakspear, M., Poldrack, R. A., Shine, J. M., & Lizier, J. T. (2019). Transitions in information processing dynamics at the whole-brain network level are driven by alterations in neural gain. *PLoS Computational Biology*, *15*(10), e1006957.

Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., & Brovelli, A. (2012). Multivoxel pattern analysis for FMRI data: a review. *Computational and mathematical methods in medicine*, *2012*.

Makovski, T., Jiang, Y. V., & Swallow, K. M. (2013). How do observer's responses affect visual long-term memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1097.

Moyal, R., & Edelman, S. (2019). Dynamic computation in visual thalamocortical networks. *Entropy*, *21*(5), 500.

Moyal, R., Fekete, T., & Edelman, S. (2020). Dynamical Emergence Theory (DET): A computational account of phenomenal consciousness. *Minds and Machines*, 1-21.

Mulligan, N. W., & Spataro, P. (2015). Divided attention can enhance early-phase memory encoding: The attentional boost effect and study trial duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1223.

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*, *59*(3), 2636-2643.

Murphy, P. R., O'connell, R. G., O'sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human brain mapping*, *35*(8), 4140-4154.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus--norepinephrine system. *Psychological bulletin*, *131*(4), 510.

Nobre, A. C., & Van Ede, F. (2018). Anticipated moments: temporal structure in attention. *Nature Reviews Neuroscience*, *19*(1), 34.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Pihlajamäki, M., O'Keefe, K., O'Brien, J., Blacker, D., & Sperling, R. A. (2011). Failure of repetition suppression and memory encoding in aging and Alzheimer's disease. *Brain imaging and behavior*, *5*(1), 36-44.

Poch, C. & Campo, P. (2012). Neocortical-hippocampal dynamics of working memory in healthy and diseased brain states based on functional connectivity. *Frontiers in human neuroscience*, *6*, 36.

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in cognitive sciences*, *17*(5), 230-240.

Ranganath, C., Heller, A., Cohen, M. X., Brozinsky, C. J., & Rissman, J. (2005). Functional connectivity with the hippocampus during successful memory formation. *Hippocampus*, *15*(8), 997-1005.

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage*, *23*(2), 752-763.

Rohenkohl, G., Coull, J. T., & Nobre, A. C. (2011). Behavioural dissociation between exogenous and endogenous temporal orienting of attention. *PLoS One*, *6*(1), e14620.

Saalmann, Y. B., & Kastner, S. (2011). Cognitive and perceptual functions of the visual thalamus. *Neuron*, *71*(2), 209-223.

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature reviews neuroscience*, *10*(3), 211-223.

Seidl, K. N., Peelen, M. V., & Kastner, S. (2012). Neural evidence for distracter suppression during visual search in real-world scenes. *Journal of Neuroscience*, *32*(34), 11812-11819.

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science*, *249*(4971), 892-895.

Shine, J. M., Aburn, M. J., Breakspear, M., & Poldrack, R. A. (2018). The modulation of neural gain facilitates a transition between functional segregation and integration in the brain. *Elife*, *7*, e31130.

Silver, M. A., Ress, D., & Heeger, D. J. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *Journal of neurophysiology*, *97*(1), 229-237.

Singer, W. (2013). Cortical dynamics revisited. *Trends in cognitive sciences*, *17*(12), 616-626.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature neuroscience*, *11*(9), 1004.

Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, *9*(3), 293-300.

Swallow, K. M., & Atir, S. (2019). The role of value in the attentional boost effect. *Quarterly Journal of Experimental Psychology*, *72*(3), 523-542.

Swallow, K. M., & Jiang, Y. V. (2010). The attentional boost effect: Transient increases in attention to one task enhance performance in a second task. *Cognition*, *115*(1), 118-132.

Swallow, K. M., & Jiang, Y. V. (2011). The role of timing in the attentional boost effect. *Attention, Perception, & Psychophysics*, *73*(2), 389-404.

Swallow, K. M., & Jiang, Y. V. (2013). Attentional load and attentional boost: A review of data and theory. *Frontiers in Psychology*, *4*, 274.

Swallow, K. M., & Jiang, Y. V. (2014). The attentional boost effect really is a boost: Evidence from a new baseline. *Attention, Perception, & Psychophysics*, *76*(5), 1298-1307.

Swallow, K. M., Jiang, Y. V., & Riley, E. B. (2019). Target detection increases pupil diameter and enhances memory for background scenes during multi-tasking. *Scientific reports*, *9*(1), 1-13.

Swallow, K. M., Makovski, T., & Jiang, Y. V. (2012). Selection of events in time enhances activity throughout early visual cortex. *Journal of Neurophysiology*, *108*(12), 3239-3252.

Turker, H. B., Riley, E., Luh, W. M., Colcombe, S. J., & Swallow, K. M. (2019). Estimates of locus coeruleus function with functional magnetic resonance imaging are influenced by localization approaches and the use of multi-echo data. *bioRxiv*, 731620.

Turker, H. B., & Swallow, K. M. (2019). Attending to behaviorally relevant moments enhances incidental relational memory. *Memory & cognition*, *47*(1), 1-16.

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic bulletin & review*, *25*(6), 2005-2015.

Wang, C. A., & Munoz, D. P. (2015). A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Current opinion in neurobiology*, *33*, 134-140.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior research methods*, *42*(3), 671-684.

Xiang, Y., Mottaghi, R., & Savarese, S. (2014, March). Beyond pascal: A benchmark for 3d object detection in the wild. In *IEEE winter conference on applications of computer vision* (pp. 75-82). IEEE.

Yebra, M., Galarza-Vallejo, A., Soto-Leon, V., Gonzalez-Rosa, J. J., de Berker, A. O., Bestmann, S., ... & Strange, B. A. (2019). Action boosts episodic memory encoding in humans via engagement of a noradrenergic system. *Nature communications*, *10*(1), 1-12.

Zweynert, S., Pade, J. P., Wüstenberg, T., Sterzer, P., Walter, H., Seidenbecher, C. I., ... & Schott, B. H. (2011). Motivational salience modulates hippocampal repetition suppression and functional connectivity in humans. *Frontiers in Human Neuroscience*, *5*, 144.

Figures

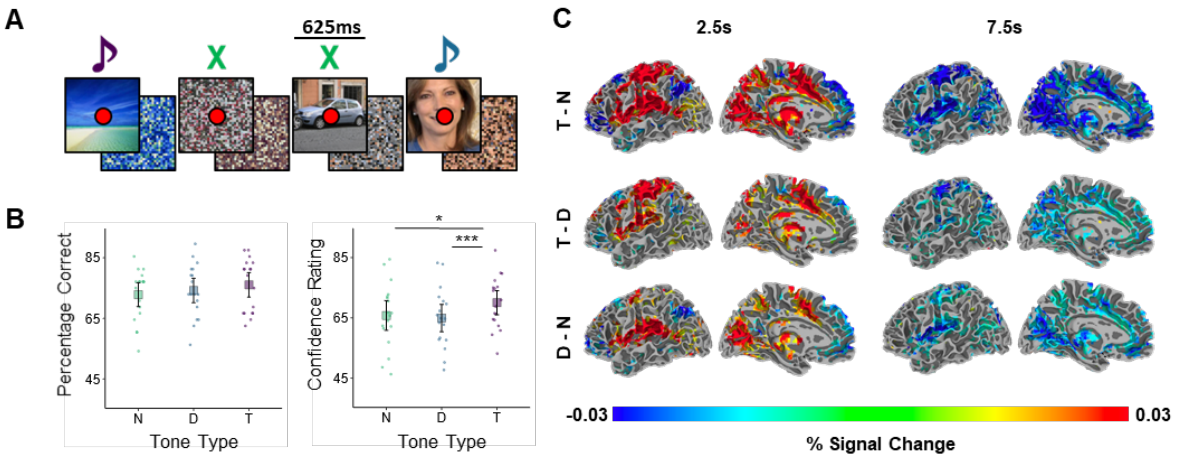

Figure 1. Experimental paradigm and the effect of tone type on subsequent memory and whole brain activity. A: The picture encoding and target detection task performed during scanning. Each stimulus was presented for 625ms. Each photograph was followed by one or more scrambled images. The photographs belonged to one of six categories and could be accompanied by a target tone (purple) that warranted a button press, a distractor tone (blue) that did not, or no tone (green). **B**: Left: Percentage of correctly recognized images during a post-scanning two alternative forced choice recognition test. Right: confidence ratings (continuous scale, 0-100) for correctly remembered images. Each point represents a single participant's mean. Large squares are centered on the sample mean and error bars indicate 95% confidence intervals around the mean. **C**: Whole-brain, group-level statistical maps illustrating the effects of tone type over time. Illustrated here are voxels showing a significant tone type by time interaction, $F > 3.483$, $q < .001$, and whose activity significantly differed across two tone conditions at 2.5 s and 7.5 s after trial onset, $z = 1.96$, $p < .05$, range of corresponding q thresholds: .016-.06. Statistical maps are overlaid on the MNI N27 atlas. Left to right: time points 2.5 s and 7.5 s. Top to bottom: target – no tone baseline (T-N), target – distractor (T-D), distractor – no tone baseline (D-N).

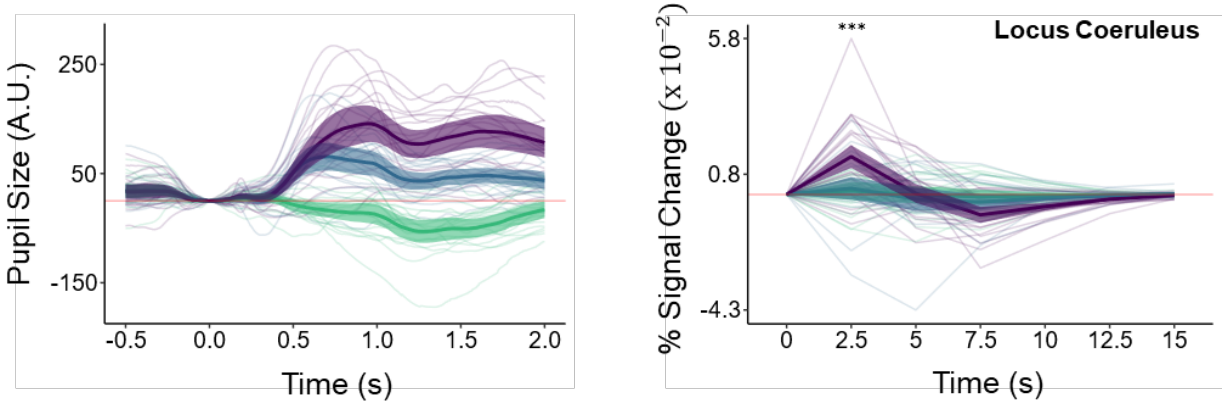

Figure 2. Auditory target detection increased phasic pupil responses and locus coeruleus activity. Left: Phasic pupil response magnitude time series during the image encoding and target detection task, as a function of subsequent image recall. Each line represents an observation from a single participant. Ribbons indicate 95% confidence intervals around the mean. Right: BOLD magnitude (% signal change) time series for the locus coeruleus (LC). Asterisks indicate that responses were significantly ($p < .001$) higher in the target (purple) compared to the distractor (blue) and the baseline (green) conditions at the respective time point.

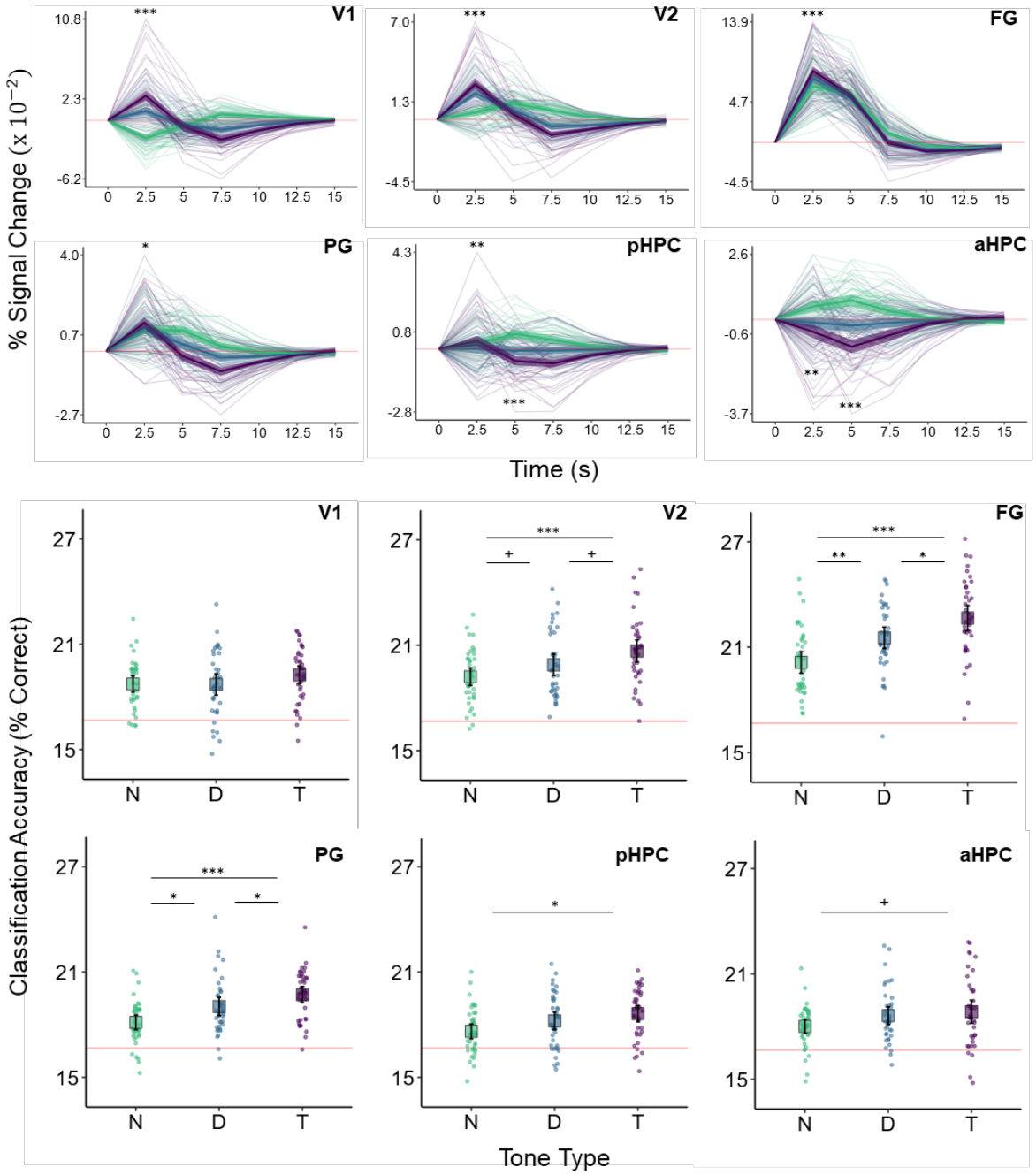

Figure 3. BOLD magnitude and image category classification accuracy as a function of tone type. Top two rows: BOLD magnitude (% signal change) time series for the visual and hippocampal ROIs in the target (purple), distractor (blue), and no tone (green) conditions. Asterisks indicating significant differences between target and distractor conditions are shown only for 2.5 s and 5 s. Faint lines show data for a single participant. Bold lines show the mean across participants. Ribbons around the thick lines

indicate the 95% confidence intervals around the mean. Bottom two rows: Six-way image category classification accuracy in the same ROIs. Theoretical chance-level performance (16.67%) is marked by a red horizontal line. Each point represents classification accuracy for a single participant. Large squares are centered on the sample mean and error bars indicate 95% confidence intervals around the mean. Asterisks denote a significant difference (p < .05*, p < .01**, p < .001***).

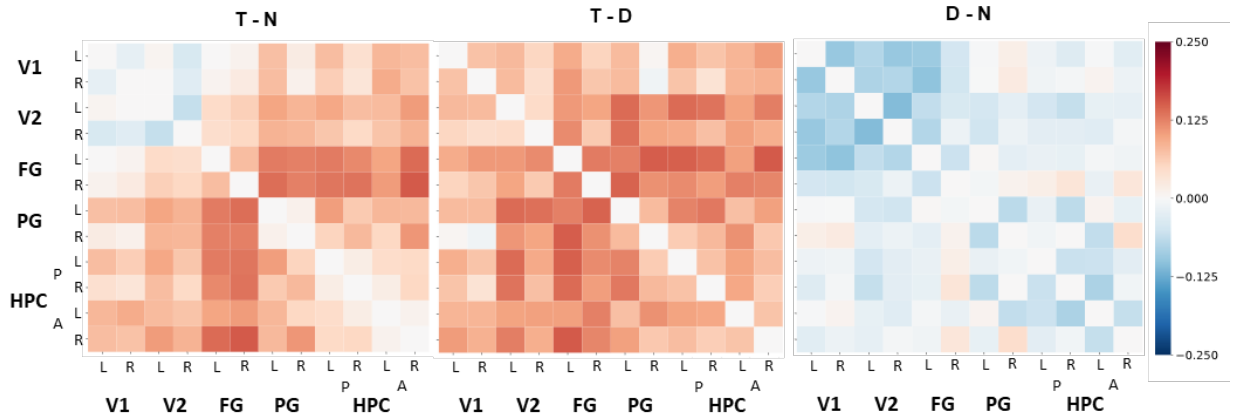

Figure 4. Tone type modulated pairwise functional connectivity throughout the visual cortex and hippocampus. Group-level pairwise functional connectivity comparisons across tone type conditions (left to right: target – no tone baseline, target – distractor, distractor – no tone baseline). Each cell corresponds to the difference between the Fisher-transformed beta series correlations of the anatomical ROIs specified in the axis labels (L = left, R = right, A = anterior, P = posterior) in the two conditions contrasted in its respective matrix.

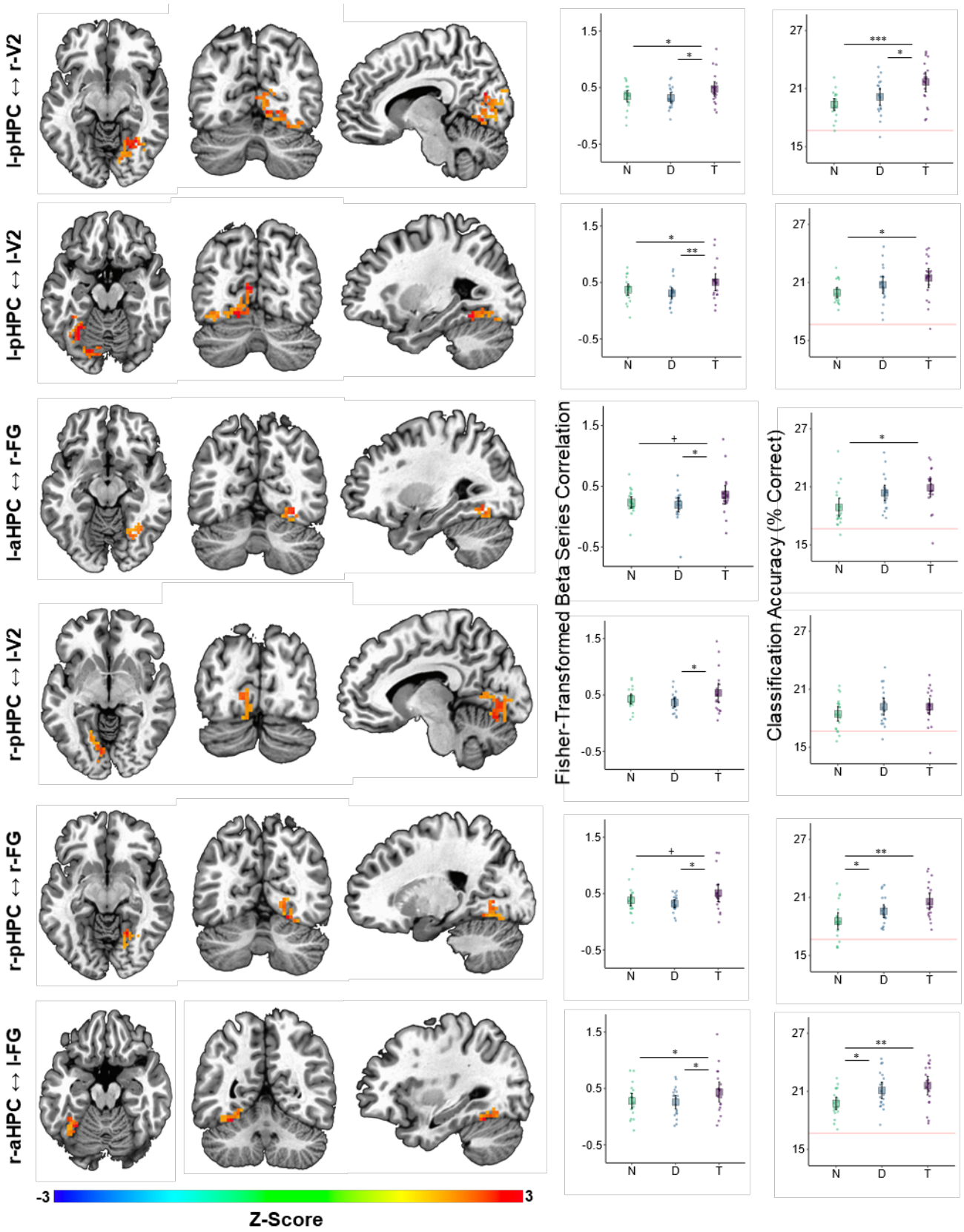
l-pHPC ↔ r-V2
l-pHPC ↔ l-V2
l-aHPC ↔ r-FG
r-pHPC ↔ l-V2
r-pHPC ↔ r-FG
r-aHPC ↔ l-FG
Fisher-Transformed Beta Series Correlation
Classification Accuracy (% Correct)
N
D
T
-3
3
Z-Score

Figure 5. Auditory target detection improved image classification accuracy in visual regions showing increased functional connectivity with the hippocampus. Left column: Clusters that were more strongly correlated with the hippocampus on target tone trials than on distractor tone trials. Voxel colors correspond to z-scores obtained by contrasting the Fisher-transformed beta series correlations between the mean beta series for the hippocampal seed region and that of each voxel in V1, V2, FG, PFG, pHPC, and aHPC. Middle column: These clusters survived a confirmatory analysis (one cluster that is not shown did not) that contrasted mean beta series correlations within each cluster in the no tone (N), distractor tone (D), and target tone (T) conditions. Each point represents the Fisher-transformed correlation coefficient for a single participant. Right column: six-way image category classification accuracy for each cluster. Theoretical chance-level performance (16.67%) is marked by a red horizontal line. Large squares are centered on the sample mean and error bars indicate 95% confidence intervals around the mean. Asterisks denote a significant difference ($p < .05$*, $p < .01$**, $p < .001$***).

Tables

				% Overlap with Anatomical ROIs							
Seed	**Cluster**	**Center of Mass**	**Size**	**l-V1**	**r-V1**	**l-V2**	**r-V2**	**l-FG**	**r-FG**	**l-PG**	**r-PG**
l-pHPC	r-V2	-16.0, 72.5, -1.1	264	0.8%	23.6%	1%	50.2%	-	24.4%	-	-
l-pHPC	l-V2	20.3, 69.0, -10.8	253	8.3%	-	48.8%	-	42.9%	-	-	-
l-aHPC	l-FG	32.4, 52.6, -17.9	60	-	-	-	-	100%	-	-	-
l-aHPC	r-FG	-30.5, 61.9, -13.9	43	-	-	-	10.5%	-	89.5%	-	-
r-pHPC	l-V2	10.8, 74.1, -5.4	102	17.5%	-	79%	-	3.5%	-	-	-
r-pHPC	r-FG	-22.9, 64.3, -10.7	56	-	1.2%	-	48.9%	-	49.9%	-	-
r-aHPC	l-FG	25.1, 66.4, -14.5	148	1%	-	37.5%	-	61.5%	-	-	-

Table 1. Visuo-hippocampal clusters. Clusters showing higher functional connectivity with the hippocampus on target tone than on distractor tone trials in the seed-to-voxel beta series correlation analysis. The percentage of voxels in a cluster that overlapped with each anatomical ROI is the mean across subjects.

ID	Center of Mass (RAI)	Size (mm^3)
1	1.3, 66.7, 12.0	837
2	-1.8, 27.9, -26.8	783
3	-28, -53.1, 17.7	729
4	28, -47.3, 26.8	675
5	.9, 42.9, 50.6	675
6	25.1, 63.1, -3.5	648
7	1.2, 80.8, 38.9	648
8	31.9, -20.1, -12.1	567
9	6.1, -33.1, 18.4	567
10	-1.1, 89,12	540
11	20.5, 41.3, -11.2	486
12	-47, 62.3, 18.2	459
13	14.1, 52.8, 15	432
14	4.3, 30.1, -9.2	405
15	-32.9, 38.8, -12.7	351
16	-4.2, 57.2, -32.7	297
17	11.6, -48.6, -7.9	297
18	5.5, 10.8, -10.7	243
19	-1.5, -9.1, -11.4	135
20	6.9, -40.9, 35.4	135

Table 2. LC-iFC clusters. RAI coordinates and sizes (mm^3) of the clusters obtained by intersecting the thresholded resting state LC iFC map, $p < .001$, with non-overlapping spherical ROIs centered on the 20 highest z-scores.