Sequential Transmission of Task-Relevant Information in Cortical Neuronal Networks

Nikolas A. Francis[1,‡], Shoutik Mukherjee[3,‡], Loren Koçillari[4,5,‡], Stefano Panzeri[4,5,*], Behtash Babadi[3,*], Patrick O. Kanold[1,2,*,†]

[1]Department of Biology & Brain and Behavior Institute, University of Maryland, College Park, MD 20742, USA

[2]Department of Biomedical Engineering & Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD 21205, USA

[3]Department of Electrical and Computer Engineering & Institute for Systems Research, University of Maryland, College Park, MD 20742, USA

[4]Laboratory of Neural Computation, Istituto Italiano di Tecnologia, Rovereto 38068, Italy

[5]Department of Excellence for Neural Information Processing, Center for Molecular Neurobiology (ZMNH), University Medical Center Hamburg-Eppendorf (UKE), D-20251 Hamburg, Germany

‡Equal co-first authorship

Keywords: auditory cortex, mouse, behavior, imaging, 2-photon, short-term memory, information, noise correlations, Granger causality, networks

Lead contact (†):

Patrick O. Kanold**,** Dept. of Biomedical Engineering, Johns Hopkins University, Miller 380. Baltimore, MD 21205 USA**.** E-mail: pkanold@jhu.edu

Corresponding authors (*):

Patrick O. Kanold**,** Dept. of Biomedical Engineering, Johns Hopkins University, Miller 380. Baltimore, MD 21205 USA**.** E-mail: pkanold@jhu.edu

Behtash Babadi, Department of Electrical & Computer Engineering, University of Maryland, College Park, MD 20742, USA. E-mail: behtash@umd.edu

Stefano Panzeri, Department of Excellence for Neural Information Processing, Center for Molecular Neurobiology (ZMNH), University Medical Center Hamburg-Eppendorf (UKE), Falkenried 94, D-20251 Hamburg, Germany. E-mail: s.panzeri@uke.de

SUMMARY

Cortical processing of task-relevant information enables recognition of behaviorally meaningful sensory events. It is unclear how task-related information is represented within cortical networks by the activity of individual neurons and their functional interactions. Here, we use 2-photon imaging to record neuronal activity from the primary auditory cortex of mice during a pure-tone discrimination task. We find that a subset of neurons transiently encode sensory information used to inform behavioral choice. Using Granger causality analysis, we show that these neurons form functional networks in which information transmits sequentially. Network structures differ for target vs non-target tones, encode behavioral choice, and differ between correct vs incorrect behavioral choices. Correct behavioral choices are associated with shorter communication timescales, larger functional correlations, and greater information redundancy. In summary, specialized neurons in primary auditory cortex integrate task-related information and form functional networks whose structures encode both sensory input and behavioral choice.

INTRODUCTION

Cortical processing of task-relevant information enables mammals to recognize behaviorally meaningful stimuli while navigating the sensory environment. Performance of an auditory task modulates neural representations of task-related sounds at the level of single neurons or small populations, already in primary auditory cortex (A1) (Kuchibhotla et al., 2017, Kato et al., 2015, David et al., 2012, Francis et al., 2018b, Tsunada et al., 2016, Brosch et al., 2011, Francis et al., 2018a, Carcea et al., 2017, Fritz et al., 2003, Insanally et al., 2019, Schwartz and David, 2018, Yin et al., 2020, Bagur et al., 2018, McGinley et al., 2015, Guo et al., 2019, Rodgers and DeWeese, 2014, Niwa et al., 2013, Christison-Lagay and Cohen, 2018). We recently showed that performing a pure-tone detection task increases neuronal responses to target tones in A1 layer 2/3 (L2/3) and changes functional connectivity by forming small strongly linked neuronal networks that encode behavioral choice (Francis et al., 2018b). However, natural auditory scenes typically include both target and non-target sounds that require discrimination. The effect of discrimination on the functional networking of neurons and how target vs non-target information propagates through the population are poorly understood.

Given the diversity of neuronal connectivity and stimulus selectivity in A1 L2/3 (Atencio and Schreiner, 2010, Atzori et al., 2001, Oviedo et al., 2010, Meng et al., 2017, Bandyopadhyay et al., 2010, Rothschild et al., 2010, Kanold et al., 2014, Maor et al., 2016, Winkowski and Kanold, 2013, Sakata and Harris, 2009, Yang et al., 2008, Sadagopan and Wang, 2009), we hypothesized that there may exist specialized neurons in A1 L2/3 that represent varying amounts of sensory or choice information and that a subset of these neurons, which carry sensory information used to inform behavioral choice, form functionally connected networks whose structural properties encode behavioral choice during task performance.

To investigate our hypotheses, we trained mice to behaviorally discriminate target vs non-target pure-tones while we recorded neuronal activity in A1 L2/3 using 2-photon (2P) Ca^{2+} imaging. We then quantified how much stimulus information (*SI*), behavioral choice information (*CI*), and intersection information (*II*), i.e., sensory information that is used to inform behavioral choice, was carried by individual neurons (Runyan et al., 2017, Panzeri et al., 2017). We used Granger causality (GC) analysis to study how these neurons were organized into functional networks (Kim et al., 2011, Sheikhattar et al., 2018, Francis et al., 2018b, Seth et al., 2015, Kaminski et al., 2001, Bressler and Seth, 2011, Quinn et al., 2015), and compiled network statistics to quantitatively compare key aspects of network structure. Here, we extended GC analysis to not only study functional network structure, but also the timescales of network interactions. Finally, to study how task-related information is transmitted and shared within functional networks, we computed information redundancy between pairs of neurons (Schneidman et al., 2003, Pola et al., 2003).

We found that task performance modulated neuronal response amplitudes, network structures, and information transmission in A1 L2/3. Individual neurons encoded *II* at different peak times which, across the population, tiled the duration of a trial. Networked neurons encoding *II* exhibited sparse connectivity and shared redundant stimulus information relevant for behavioral choice. Network structures differed for target vs non-target tones, encoded behavioral choice, and differed between correct and incorrect behavioral choices. Moreover, choice-dependent functional networks also showed different communication timescales. Together, our results describe how networked neurons in A1 L2/3 that integrate sensory and behavioral information during auditory task performance sequentially transmit task-related information.

RESULTS

To study how task-relevant information is transmitted within neuronal networks, we trained 9 transgenic CBA x Thy1-GCaMP6s F1 mice (Frisina et al., 2011, Dana et al., 2014) to perform a pure-tone frequency discrimination task (Kuchibhotla et al., 2017, Pi et al., 2013) while we imaged neuronal responses in A1 L2/3 using *in vivo* 2P Ca^{2+} imaging (Fig. 1).

Head-fixed mice learned to perform an auditory tone discrimination task

Head-fixed mice were trained to lick a waterspout in response to hearing a low-frequency target tone (Fig. 1A; 7 or 9.9 kHz, red), and to avoid licking the waterspout after hearing a high-frequency non-target tone (14 or 19.8 kHz, blue). The four frequencies were randomly interleaved across trials. Fig. 1B shows that the mice learned to behaviorally discriminate targets vs non-targets. Each trial’s behavioral response was categorized into four groups, based on the first lick on each trial: hit (H: licking after target onset), miss (M: no licking after a target), false alarm (F: licking after non-target onset), or correct rejection (C: no licking after a non-target). The middle panel of Fig. 1B shows the distribution of H and F behavioral response times (i.e., the time of the first lick in a trial). The average H and F response latencies relative to stimulus onset were 0.64s ± 0.02s and 0.75s ± 0.04s, respectively. Across the 34 experiments, the hit rate (78.8% ± 5.1%) was significantly higher than the false alarm rate (F = 27.1% ± 7.3%; $p<0.001$, t-test) and the correct rejection rate (74.3% ± 6.9%) was significantly higher ($p<0.001$, t-test) than both the F and M rate (20.8% ± 5.2%). Thus, the mice were able to discriminate between target vs non-target tones (d' = 1.4 ± 0.4).

Decision-making modulated neuronal response amplitude in A1 L2/3

To characterize neural responses during behavior, we imaged Ca^{2+}-dependent fluorescence in auditory cortex. To localize 2P imaging fields for each experiment to A1, we first mapped the tonotopy of the auditory cortex in each mouse using widefield imaging (Fig. 1C) (Liu et al., 2019, Francis et al., 2018b).

We performed 2P imaging (Fig. 1D-G) at a depth of 150-250 µm from the cortical surface in each mouse (34 experiments, 9 mice, N = 2792 neurons). We observed fluorescence ($\Delta F/F$) responses to all 4 tones with response dynamics typical of GCaMP6s (Chen et al., 2013, Dana et al., 2014). Similar to previous studies, neural traces showed a complex pattern of task-dependent changes in response amplitude (Bagur et al., 2018, Brosch et al., 2011, Carcea et al., 2017, David et al., 2012, Francis et al., 2018a, Francis et al., 2018b, Guo et al., 2019, Kato et al., 2015, Kuchibhotla et al., 2017, Rodgers and DeWeese, 2014, Schwartz and David, 2018, Tsunada et al., 2016, Yin et al., 2020).

To make sure our results do not reflect neural coding of task-related movement (i.e., licking/behavioral choice) or reward, we first quantified how neuronal response amplitude varied with task performance and pure-tone frequency during passive trials, in which the mouse sat quiescently hearing tones without doing an auditory task, and then during behavioral task performance. To make a fair comparison, we averaged neural traces only during the first 0.5 s after tone onset for the trials with behavioral response latencies greater than 0.5 s after stimulus onset. This accounted for 74% of H trials. While licking has been shown to impact activity in auditory cortex (Nelson and Mooney, 2016), we previously showed that it did not drive neural activity in A1 L2/3 during a go/no-go task (Francis et al., 2018b).

As shown in Fig. 1F, we found significant neural responses ($p<0.001$) at all frequencies in both passive and behavior conditions, indicating that the mice could hear the tones. This result, in combination with our finding that behavioral responses were similarly time-locked to both low and high-frequency tone presentations in our go/no-go task (Fig. 1B), provides compelling evidence that the mice were in fact doing a target vs non-target discrimination task, and not simply target detection. Overall, responses to non-target tones were smaller than those to target tones ($p<0.001$). In contrast, trials without behavioral responses (M & C) had the lowest average response amplitudes and there were no significant differences in neuronal responses to target vs non-target frequencies ($p>0.05$, Kruskal-Wallis test). Thus, the amplitude of pure-tone responses in A1 during task performance was strongly modulated not only by acoustic stimulation, but also by behavioral choice.

Aberrant attentional gain in A1 L2/3 reflects incorrect decision-making

The change in neuronal response amplitude to the same sound for passive vs behavior trials quantifies attentional gain in A1. As shown in Fig. 1G, correct behavioral choices (i.e., hits and correct rejections) had a small, but significant negative gain (7 kHz, H: -1.9% +/- 0.97%, $p<0.001$; 9.9 kHz, H: -1.2% +/- 1.01%, $p<0.05$; 14 kHz, C: -1.85% +/- 1.08%, $p<0.001$; 19.8 kHz, C: -0.5% +/- 1.02%, $p>0.05$). In contrast, attentional gain for incorrect behavioral choices was more varied. False alarms occurred when neuronal responses had a small positive gain (14 kHz, F: 1.97% +/- 1.23%, $p<0.001$; 19.8 kHz, C: 1.96% +/- 1.42%, $p<0.01$), whereas misses occurred when responses had a large negative attentional gain (7 kHz, M: -9.76% +/- 1.25%, p<0.001; 9.9 kHz, C: -7.7% +/- 1.24%, p<0.001). Thus, incorrect decision making was associated with aberrant attentional gain, i.e., deviations from the slight negativity observed during correct decision making.

Task-relevant information is transiently encoded by individual neurons in A1 L2/3

We hypothesized that single neurons in A1 might represent varying amounts of sensory or choice information. We first performed spike inference, as summarized in Fig. 2A and described in detail in Methods, followed by quantifying the task-relevant information carried by each neuron in single trials using information theory (Shannon, 1948, Quian Quiroga and Panzeri, 2009). For each neuron we quantified how much information was present about the acoustic stimulus (*SI;* i.e., target vs non-target tone; Fig. 2B, left) and about the behavioral choice (*CI*; i.e., lick vs no-lick; Fig. 2B, right). We also computed intersection information, (*II*; Fig. 2B, middle) (Panzeri et al., 2017, Pica et al., 2017), which quantifies how much of the sensory information encoded by the neurons is used to inform behavioral choices, and is thus a direct measure of task-relevant information. We found that 1183/2792 neurons (42%) carried either significant *SI* or *CI* (permutation test, p<0.1, corrected for comparisons across multiple time windows, see Methods). 708/1183 neurons did not carry significant *II* (permutation test, p>0.1), i.e., they either had stimulus information that did not inform choice (e.g., stimulus response was not causal to formation of the choice), or choice information not related to the stimulus (e.g., internal choice bias). The remaining 475/1183 neurons carried significant *II*, *SI,* and *CI* (permutation test, p<0.1), and thus integrate both sensory and behavioral information that is directly relevant for the decision-making task (Panzeri et al., 2017).

To better illustrate the dynamics of the information carried by either the whole population or by the neurons with significant *II*, *SI* and *CI* (shortened to “*II* neurons” hereafter), we computed the average *II*, *SI,* and *CI* time-courses. *II* neurons carried the largest amounts of *II, SI* and *CI* throughout the trial (Fig. 2C). Moreover, *CI* was highly correlated to *SI* (Pearson correlation = 0.9) for *II* neurons, while this correlation was smaller across the entire population (Pearson correlation = 0.51), suggesting that similar levels of *CI* and *SI* are present across *II* neurons (Fig. 2C). At the time of peak information, each neuron carried 0.117 ± 0.003 bits of *SI*, 0.121 ± 0.004 bits of *CI* and 0.085 ± 0.002 bits of *II*, respectively (mean ± SEM across the 475 *II* neurons).

We restricted our further analyses to 12 out of 34 sessions with at least 20 *II* neurons due to our interest in subsequent network analyses for which, given the number of experimental trials, up to 20 neurons could be analyzed with statistical confidence. This left us with 375/475 *II* neurons for subsequent analyses (see Methods). Figure 3A shows the peak-normalized information time-courses for each *II* neuron, sorted by the peak-latency of the respective information (*SI, CI* or *II*) across neurons. Qualitative inspection showed that neurons transiently encoded *SI*, *CI*, or *II*, tiling the trial duration. The *SI, CI,* and *II* time-courses of *II* neurons showed similar average trends, while neurons that carried either significant *SI* or *CI,* but not *II*, showed more heterogeneous trends (Supp. Fig. S1).

We next computed the *II*/*SI* and *II/CI* ratios to identify how much of the stimulus and choice information was used for informing behavioral choice. The average *II/SI* and II/CI ratio was high (>70%) for *II* neurons throughout the trial, meaning that most of the *SI* was used for informing choice and most of the *CI* reflected stimulus discrimination rather than a stimulus-unrelated choice bias (Fig. 3A). Conversely, neurons with significant *SI* or *CI,* but not *II,* showed an overall decrease of information within the 500 ms waiting period after tone onset (Supp. Fig. S1). In addition, they showed lower *II/SI* or *II/CI* ratios down to ~30%, especially at the peak times, meaning that *SI* and *CI* are not optimally used to perform the task. These results indicate that we identified neurons in A1 L2/3 that transiently carry significant stimulus information used to inform behavioral choice. To quantify this transiency, we aligned information peaks across neurons and analyzed the peak-aligned traces within ± 1 s of the peak (Fig. 3A), which admitted an exponential fit with a time constant $\tau \sim \frac{1}{3.9}s \sim 250\ ms$. Thus, individual neurons transiently carried *SI, CI, II* for an effective duration of $\sim 250\ ms$.

To inspect the dynamics of information carried across the trial duration, we clustered the neurons based on their *II*-peak latencies. We labelled neurons that peaked in the first 1.5 seconds after stimulus onset as *peri-stimulus II*, and the remainder as *post-stimulus II* (Fig. 3B). We subdivided the *peri-stimulus II* neurons into three sequential task-related periods within a trial: (1) the 500 ms waiting period just after tone onset, (2) the 500 ms interval after the waiting period, and (3) the 500 ms after tone offset (labeled respectively in Fig. 3B, left column). We found that 52/375, 85/375, and 60/375 neurons had *II* that peaked in the first, second, and third *peri-stimulus* periods, respectively (Fig. 3B, left column), adding up to 197/375 neurons. Furthermore, 45/375, 40/375, and 48/375 neurons had *II* that peaked in the fourth, fifth, and sixth *post-stimulus* periods (1.5-3 s), respectively (labeled respectively in Fig. 3B, right column). The remaining 45/375 neurons peaked after 3 s. Although the values of *SI*, *CI* and *II* remained comparable throughout the trial, neurons with earlier responses carried slightly more *SI* than *CI* (Fig. 3B, left column, blue vs green traces) and neurons with later responses carried slightly higher *CI* than *SI* (Fig. 3B, right column, green vs blue traces).

Given that *II* neurons carried *SI,* we next examined their tuning properties (Fig. 3C). We found that the best frequencies (BFs; the frequency values eliciting the highest response during passive tone presentation) of *II* neurons were lower ($p<0.01$, Wilcoxon rank sum test) than the average BF of the overall population. *II* neurons also had narrower bandwidth (BW) ($p<0.05$, Wilcoxon rank sum test) than the overall population. Our results show that task-relevant information was transiently encoded by individual neurons, yet sustained throughout the trial by sequential encoding across a population of neurons in A1 L2/3.

Neurons with intersection information form sparse timescale-invariant functional networks

Since individual neurons had low information content and only transiently encoded *II* (Fig. 2B,C; Fig. 3A), we hypothesized that *II* neurons might form functional networks to more robustly encode task-relevant information. We previously used Granger causality (GC) analysis to identify small functional networks of interacting neurons whose network structure depended on behavioral choice (Francis et al., 2018b), but did not study how network structure might vary with integration timescales, i.e., the duration over which neurons might interact. Our finding that *II* was transiently encoded by individual neurons, but sustained across time by the population, necessitated the examination of relevant timescales of interactions between *II* neurons. Hence, we extended our previous GC analysis by considering the interaction timescales.

GC analysis uses multivariate statistics to infer causal influences within a population of neurons by testing if the recent history of a neuron can improve the prediction of another neuron's activity. The duration of the recent history over which interactions are quantified, referred to as the ‘integration window’, is a hyperparameter of GC analysis, whose value, *w*, sets the longest interaction window considered (Fig. 4A, left schematic). Short (S; *w*=233 ms) integration windows quantify dynamics that are more likely to reflect local neuronal interactions. Long (L; *w*=1033 ms) integration windows would additionally capture the effects of potentially slower and indirectly mediated interactions that may involve distant neurons. The specific values of *w* we used were integer multiples of the imaging frame rate. Importantly, the S-timescale interactions are a subset of the L-timescale interactions (see Methods). For each experiment (N=12), we performed GC analysis on the 20 neurons with the lowest, i.e, “shortest”, *II*-peak latencies to identify the contribution of neurons whose activity carried task-relevant information during stimulus presentation. We used 20 neurons per experiment to avoid overfitting the data, given the limited number of experimental trials (see Methods). GC networks were estimated individually for each behavioral choice category (H, M, C, and F) in the *discrimination* task, importantly contrasting previous work (Francis et al., 2018b) in which we could only analyze networks corresponding to H and M categories in a *detection* task.

We found that across all trials GC networks were sparse: only 1% of possible links connecting 21.98% of the selected *II* neurons were detected in S-timescale networks, while 3.61% were detected in L-timescale networks connecting 51.67 % of the selected *II* neurons. Unlike simpler measures such as Pearson correlation, GC is a directed measure of communication, which can distinguish senders from receivers (Figure 4A). This allowed us to investigate the proportion of senders and receivers within the network. For the S-timescale networks, 10.10% of neurons were senders, 9.06% were receivers, and 2.81% were GC-linked neurons that had net degree of zero. For the L-timescale networks, however, 24.58% of the neurons were senders, 19.79% were receivers and 7.29% had a net degree of zero. This indicates that an additional 29.69% of the selected *II* neurons were recruited over the longer integration window.

Speculating that the information content of GC-linked neurons differed from GC-unlinked neurons, we compared *SI, CI,* and *II* at the *II* peak time*.* We found that *SI, CI,* and *II* were higher in GC-linked than in GC-unlinked neurons, in both S- and L-timescale networks (Figure 4B). These results suggest that GC-linked neurons form networks carrying signals of greater relevance for performing the auditory discrimination task. Given that neurons carry information over a wide range of timescales (Fig. 4B), we tested whether neurons' *II*-peak latencies depended on their membership in S- or L-timescale networks. We compared the distributions of *II*-peak latencies of neurons in S- vs L-timescale networks and found no significant difference (Wilcoxon rank sum test, p=0.5847). Furthermore, we compared the *II*-peak latency for pairs of GC-linked neurons and found that the *II*-peak of receiver follows that of sender neurons by 2.07 time bins (69 ms) on average; however this difference was not statistically significant (Wilcoxon signed rank test, p=0.1743). As such, we found no evidence to support that *II*-peak latency correlates with membership of the linked neurons in S- vs L-timescale networks or the direction of the GC links.

To characterize how the structure of the GC networks depends on the timescale of interactions and on behavioral choice, we analyzed 4 network statistics separately for H, M, C, and F trials: number of links, number of subnetworks (isolated subsets of neurons), subnetwork size (number of member neurons), and statistical strength of links (Youden's *J*-statistic) (Francis et al., 2018b) (shown from left to right in Fig. 4C; see also Supp. Table S1). We focused on neurons with low *II*-peak latency, as they are more likely to be causally related to choice. For both M and F networks (incorrect behavior), the number of links and the size of subnetworks were greater for L- than S-timescale networks, while link strength was less for L- than S-timescale networks. In contrast, we found no differences for L- vs S-timescale networks in H or C trials (correct behavior), for the number of links, size of subnetworks, and link strengths. In C trials, number of subnetworks increased with integration window length. Together, our results show that incorrect decision (M & F) L-timescale networks are larger but connected less strongly than their S-timescale counterparts. In contrast, the structure of correct decision networks (H & C) was invariant across timescales. Noting that S-timescale interactions are a subset of the L-timescale ones in our model, the invariance of the correct decision network structure between S- and L-timescales may suggest the involvement of a network of local cortical interactions, rather than of interactions mediated by wider loops involving farther neurons.

For S-timescale, comparison of H and M networks showed the former had more links and larger subnetworks, suggesting that larger networks are beneficial for encoding correct detection of the target. The average link strength was greater in C than in F networks, suggesting that stronger links are beneficial for encoding correct rejection of the non-target. In contrast, for L-timescale, both the number of links and the sizes of subnetworks were smaller for correct than incorrect categories, while links remained stronger for the correct category.

Neuronal network structure encodes behavioral choice

Since the GC network structures for neurons with low *II*-peak latency strongly depended on behavioral choice, we sought to directly test if the network structures encode behavioral choice. Thus, we used the 4 network statistics as features for a support vector machine (SVM) trained to distinguish between correct (H & C) and incorrect (M & F) decisions. For comparison, we trained a similar classifier for networks of neurons with high response rates, chosen regardless of the information content they carry. The comparison between the network structure of *II* neurons and responsive neurons is non-trivial because a pair of neurons both carrying *II* is neither necessary nor sufficient for there to be a GC link between them (See Supp. Fig. S2 for counterexamples). The GC network statistics of highly responsive neurons are reported in Supp. Table S2. Of all low *II*-peak latency neurons, 30.21% were also identified as highly responsive neurons (see Methods for selection criterion). The network structure of S-timescale networks for low *II*-peak latency neurons classified behavioral choice much more accurately than that of highly responsive neurons (Fig. 4D, left bar plots). In contrast, the features of L-timescale networks classified behavioral choice well for both low *II*-peak latency and highly responsive neurons, though more accurately for the former (Fig. 4D, right bar plots). These results show that S-timescale networks of low *II*-peak latency neurons better encode behavioral choice than those of highly responsive neurons and suggest that low *II*-peak latency neurons form a specialized group of neurons in A1.

One possibility is that strong choice predictivity from network interactions is not a special property of networks formed by *II* neurons, but is also present in networks of neurons with either *SI* not used for choice or *CI* not related to the stimulus. To test this possibility, we compared the predictivity of low *II*-peak latency neurons to that of *SI* and *CI* neurons that did not have significant *II* (Supp. Fig. S3A-B and Supp. Table S3). While 7 sessions had a sufficient number of exclusively *CI* neurons, only 1 session had a sufficient number of exclusively *SI* neurons for GC network analysis; hence, we focused on the networks of exclusively *CI* neurons to contrast with the low-latency *II* neurons analyzed in Fig. 4. Network structures of low *II*-peak latency neurons were more predictive than *CI* neurons (Supp. Fig. S3B). Furthermore, the network structures of neurons with the greatest *II*-peak magnitudes (Supp. Fig. S3C-D and Supp. Table S4) were also more predictive of behavioral choice than highly responsive neurons (Supp. Fig. S3D). Our results suggest that the encoding of behavioral choice in the S-timescale network structure is specific to *II* neurons, and it is not found as much in groups of neurons with choice information not related to the stimulus.

The spatial extent of neuronal subnetworks varies less by timescale during correct behavioral choices

Since 2P imaging gives the exact spatial location of each neuron in a field of view, we sought to characterize how *II* neurons and their functional networks were distributed spatially. We first studied if neurons with *II* peaks in *peri* vs *post-stimulus* intervals were in different regions or if they were intermingled. We calculated the sum of the average distances of *peri- (Pe)* and *post-stimulus (Po)* neurons to their centroids (R_{Pe} and R_{Po}, respectively) and compared the sum to the distance between the centroids ($R_{Pe\,-\,Po}$). The distance between centroids was smaller than the spread of each set of neurons (Fig. 5A). Thus, *Pe* and *Po* neurons were heterogeneously distributed within the field of view, suggesting that information flow did not have intrinsic directionality from one subarea to another during task performance.

We next analyzed how subnetworks were dispersed by computing the vector distances of subnetworked neurons to the subnetwork centroid (Fig. 5B, top schematic; see also Methods). Subnetworks of L-timescale interactions tended to be more spatially dispersed than the S-timescale ones (Fig. 5B, bottom subpanels), as indicated by the determinant of the distance vector covariance matrix (Fig. 5C). The dispersion of M, F, and C subnetworks were larger for L- than S-timescales. Differences in dispersion were also observed between H and M trials for S-timescale subnetworks and between C and F trials for L-timescale subnetworks. To see if differences in the dispersion of subnetworks across timescales were due to greater distances between linked neurons, rather than the inclusion of additional neurons, we computed the average pairwise distance between linked neurons, i.e., the average link length (Fig. 5D). Except for M networks, GC link lengths were stable across timescales, indicating that the greater subnetwork spatial dispersion for L-timescale interactions was more likely due to the inclusion of additional neurons than an increased distance between linked neurons. These results suggest that correct choices are associated with spatially stable compact subnetworks while incorrect choices involve activity spread to additional neurons.

Networked neurons communicate task-relevant stimulus information that reverberates redundantly

A functional link between neurons suggests that task-relevant information is transmitted from one neuron to another. This would create a population code whose information content is reverberated redundantly across neurons because the same information is shared by different neurons.

To investigate the nature of information present in the functional networks, we measured information redundancy (Schneidman et al., 2003, Pola et al., 2003) between GC-linked neurons (Fig. 6A). We used a normalized redundancy index defined as the information carried jointly by two neurons minus the sum of the information that each carried independently, normalized with respect to the total information carried by the two neurons jointly. The value of the normalized redundancy index indicates the fraction of total joint information that is shared by two neurons. Neurons share redundant information when the redundancy index is negative, i.e., together they carry less information than the sum of the information they carry separately. Positive values of the redundancy index are associated to synergy, i.e., the contribution from the interaction between the neurons to the joint information cannot be inferred by considering each neuron individually.
For *SI, CI,* and *II*, we computed redundancy at the peak time of *II*, for each pair of neurons used in GC analysis. Normalized redundancy between pairs of neurons with a S-timescale GC link was compared to those with no GC link (Fig. 6A). For S-timescales, we found that information shared by pairs of neurons was redundant. This implies that neurons shared part of the information they

transiently carried at different times. Normalized redundancy was much larger for *II* than *CI* and *SI* (Fig. 6A, left bar plots vs middle and right bar plots). For L-timescales, we found similar trends but proportionally smaller variations between GC-linked and GC-unlinked neurons (Supp. Fig. S4). This means that neurons shared more of the behaviorally-relevant, than behaviorally-irrelevant, portion of the *SI* they carried. Importantly, the difference between normalized redundancy for GC-linked vs GC-unlinked neurons was much larger for *II* than for *SI or CI* (Fig. 6A, red vs black bar plots), reinforcing the interpretation that S-timescale GC links mediate the exchange of behaviorally-relevant sensory information.

GC links can be positive or negative valued, reflecting functionally facilitative or suppressive interactions, respectively (Francis et al., 2018b, Sheikhattar et al., 2018). We found that negative GC links had a much larger effect on redundancy, suggesting they mediate more *II* exchange than positive links (Fig. 6A, orange vs yellow bar plots). These results might indicate a mediating role of inhibitory circuits in task-related network activity (Kuchibhotla et al., 2017). Sorting the normalized redundancy with respect to the *II* peak time lags (Fig. 6B, left panel) revealed that *II* redundancy varies across time lags with an overall increasing trend (from -0.1 to -0.15). This indicates that redundant information persists during the trial.

Previous studies showed that nearby cells typically interact redundantly (Nirenberg et al., 2001, Reich et al., 2001, Chechik et al., 2006). We thus investigated how redundant information spreads spatially for *II, SI,* and, *CI* by plotting the time-lagged redundancy as a function of the Euclidean distance between pairs of neurons (Fig. 6B, right panel). We found a peak of redundant interaction for *II* at a distance of $\sim 50\ \mu m$ $(II = -0.1971 \pm 0.0187)$ which then reached a plateau at $\sim 320\ \mu m$ $(II = -0.1499 \pm 0.0039)$, followed by a distance-independent trend. *SI* and *CI* were similarly redundant and reached a plateau at $\sim 208\ \mu m$ $(SI = -0.0691 \pm 0.0019, CI = -0.0709 \pm 0.0019$. Together, these results suggest that GC links indicate redundant communication of behaviorally-relevant stimulus information, and that redundant neurons are located in close proximity of each other.

Signal correlations, noise correlations and redundancy in correct vs incorrect trials

Redundancy is critically shaped by signal and noise correlation (Schneidman et al., 2003, Pola et al., 2003). To gain more insights into the origin of redundancy, we quantified *noise correlations*, i.e., the single-trial covariations of activity that are unrelated to stimulus signals, and *signal correlations*, i.e., the covariations of activity that reflect similarity in trial-averaged stimulus tuning (see Methods).

Fig. 6C shows the signal and noise correlations for GC-linked and unliked pairs of neurons computed at the *II* peak times using all available trials. Noise correlations reduce population information (and thus contribute to redundancy) when they have the same sign as the signal correlations (Schneidman et al., 2003, Pola et al., 2003). In our analysis, both signal and noise correlations were on average positive (Fig. 6C). Thus, the observed redundancy reflects the matching positive signs of both signal and noise correlations. GC-linked pairs of neurons exhibited higher noise correlations than GC-unlinked pairs of neurons, consistent with the view that trial-to-trial correlations should be stronger for neurons that are functionally linked. Signal correlations did not vary ($p>0.05$, t-test) between GC-linked and GC-unlinked pairs of neurons.

To examine the possible advantages of the observed correlations and redundancy values for task performance and behavioral accuracy, we next refined them by separating correct (H or C) and incorrect (M or F) trials (Fig. 6D). We found several results of interest. First, in correct trials, noise correlations were stronger and had the same sign as the signal correlations (Fig. 6D, two leftmost subplots), leading to a strong redundancy of information in correct trials (Fig. 6D, third subplot).

Second, neurons did not exhibit redundancy in incorrect trials, and were indeed *synergistic*, i.e., with positive redundancy index (Fig. 6D, third subplot). This property held for both pairs that were or were not GC-linked. It is important here to note that GC-linked neurons had stronger noise correlations in correct trials (Fig. 6D, leftmost subplot), suggesting that stronger noise correlations during correct behavior results, at least in part, from network communication as revealed by GC analysis. The source of the synergy between neurons during incorrect trials, despite having positive signal and noise correlations, is further examined in Supp. Fig. S5.

Third, when computing the difference between the normalized redundancy index in correct and incorrect trials, we observed that there was more redundancy in correct trials for GC-linked pairs (Fig. 6D, rightmost panel). Together, these results suggest that redundancy and noise correlations may provide advantages for task performance. In addition, greater redundancy during correct behavior choices may partly result from increased within-network communication.

DISCUSSION

In this study we found that during the performance of an auditory discrimination task, individual neurons in A1 L2/3 transiently carried information about the stimulus (*SI*), behavioral choice (*CI*), or both (*II*) for hundreds of milliseconds, and that task-relevant information was sustained across the duration of a three-second trial by sequential propagation of *SI, CI,* and *II* in functionally connected neuronal populations. Furthermore, we identified a subpopulation of low *II*-peak latency neurons, which formed functionally connected networks whose structure could reliably predict behavioral choice. Our findings suggest that the spatiotemporal structure of functional connectivity between low *II*-peak latency neurons in A1 L2/3 may form a neural population base for sustained representation of task-relevant information.

A1 L2/3 contains a diverse population of neurons with differing functional connectivity (Meng et al., 2017, Liu and Kanold, 2021). We here find that the bandwidth of *II* neurons is lower than that of other neurons. This suggests that these neurons might be part of a class of A1 L2/3 neurons which receive L4 inputs and have limited integration across the tonotopic axis (Meng et al., 2017).

Task Relevance of Short (S) vs Long (L) Timescale Neuronal Interactions

The nested parameterization of L- and S-timescale interactions allowed us to differentiate between solely S-timescale vs additional L-timescale interactions in functional networks. Comparing L- vs S-timescale networks showed that correct choice L-timescale networks consisted of fewer but stronger links that were mostly S-timescale influences. In contrast, incorrect choice networks are characterized by a mixture of both S- and L-timescale links, and by an increased network size due to recruitment of additional spatially distant neurons nearby (within a 2P field of view). These additional L-timescale links likely reflect the local effects of slower interactions with distant neurons, perhaps reflecting non-sensory task-related interaction, such as error-signaling or deviance detection (Parras et al., 2021, Steinmetz et al., 2019, Stringer et al., 2019, Chen et al., 2015, Khouri and Nelken, 2015). Since subnetworks during correct trials varied less by timescale, this suggests that the influence of more distant neurons is suppressed when correct decisions are made, leaving predominantly S-timescale interactions. Such suppression could be mediated by inputs to A1 which can activate inhibitory circuits (Fritz et al., 2010, Winkowski et al., 2018, Winkowski et al., 2013, Liu et al., 2021).

Magnitude of stimulus, choice and intersection information in A1 L2/3 neurons

We used information theory to characterize neural selectivity to the task variables, and quantify, as function of trial time, the amount of *SI, CI,* and *II*. Our non-parametric approach for establishing neural selectivity (Quiroga and Panzeri, 2009) is relatively assumption-free and can capture linear dependencies on stimulus and choice (which are commonly captured using parametric approaches such as Generalized Linear Models) as well as possible non-linear interactions between stimulus and choice. The latter non-linearities are expected to exist under general conditions (Chicharro et al., 2021) and are difficult to reveal with parametric approaches unless one introduces strong model assumptions.

The amounts of *SI* and *CI* per neuron were similar, and approximately of 0.1 bits at the information peak. Given that 1 bit of information is needed to solve the binary discrimination task and neurons were found to carry partly redundant information, a first implication of these values is that task performance must rely on networks of at least tens of neurons. This consideration is compatible with the high behavioral choice prediction afforded by networks of 20 functionally connected neurons. Previous studies of A1 L2/3 reported much higher values of *SI* than *CI* (Runyan et al., 2017, Pica et al., 2017). We interpret these dissimilarities as largely influenced by task design, which is distinct from the one used in Runyan et al. (2017), rather than a difference of computations. In contrast to our use of a go/no go task in tone discrimination, these previous studies involved discrimination of the location of broad-band sounds during spatial navigation in a virtual reality setting, with large delays between stimulus and reward, as well as using a forced-choice task, in which all stimuli could be associated with a reward. Here, we found values of *II* to be close to those of *SI*, suggesting that most of the auditory information carried by the neurons we identified was used to inform choice, supporting the assumption that the identified functional networks are important for the execution of the task.

Population Coding via Reverberation of Redundant Information in Networks

We found high redundancy between the behaviorally-relevant stimulus information carried at the time of information peaks between pairs of low *II*-peak latency neurons. The redundancy was higher between pairs of neurons that were GC-linked based on S-timescale interactions, suggesting that the GC link may reflect the transfer of behaviorally-relevant information from one neuron to another. Redundancy has been traditionally viewed as a negative feature of population coding that should be reduced, based on theories of efficient coding (Attneave, 1954, Barlow, 1961, Nigam et al., 2019), and on the often implicit assumption of optimal information readout, implying that higher neural information corresponds to better performance (Gold and Shadlen, 2001). However, other studies have proposed that high values of spatiotemporal redundancy might facilitate biophysical signal propagation (Valente et al., 2021, Salinas and Sejnowski, 2001, Alonso et al., 1996).

Recent studies have proposed that the benefits of redundancy may outweigh its disadvantages by making task-related signals available for longer timescales (Runyan et al., 2017) and facilitating the behavioral readout of the sensory signal (Valente et al., 2021). Here, our results of higher redundancy and stronger noise correlations during correct behavior, especially for GC-linked neurons, confirm the prediction of these previous works and add the new insight that stronger noise correlations and higher redundancy during correct behavior may partly originate from increased within-network communication as revealed by GC analysis.

Importantly, previous studies highlighting the role of redundancy in behavioral readout (Runyan et al., 2017, Valente et al., 2021) concentrated only on the average strength of pairwise noise correlations. In contrast, we characterized the network-level structure of behaviorally-relevant information sharing and of correct perceptual decisions. We found that higher redundancy in GC-linked neurons was accompanied by a higher number of links, larger subnetworks in correct target detection, and stronger links in correct rejection of non-targets. Together, redundancy and GC analyses suggest that correct decisions in an auditory discrimination task may require temporary reverberation of information in the spatiotemporal structure of neuronal networks. This might explain larger redundancy for behaviorally-relevant than behaviorally-irrelevant sensory information.

In summary, our results show that during behavioral discrimination of pure-tone frequencies, task-relevant information is transmitted sequentially across individual neurons in A1 and is sustained for long periods of time within compact neuronal networks.

Limitations of the study

Our study demonstrates a statistical relationship between the considered features of neural activity and behavior, but does not causally demonstrate the role of these features in generating correct and incorrect perceptions. Only perturbations of and observation of the behavioral changes that manipulations of such features (e.g., using holographic 2P optogenetics) induce could establish causality. Furthermore, we found similarities and differences with previous studies of the behavioral relevance of emergent features (e.g., correlations, redundancy, GC networks) of population codes in the auditory cortex. We speculated that some of these differences may be due to task differences. Since neither previous studies nor ours systematically manipulated the behavioral task design, we could not establish how our conclusions would vary in different task conditions.

ACKNOWLEDGEMENTS

The authors thank Menatallah Mohamed for surgical assistance and Marco Celotto for providing software. Supported by NIH RO1DC9607 (POK), U01NS90569 (POK), P01AG55365 (POK), U19NS107474 (POK, BB, SP), R01DC17785 (POK), R21DC017829 (NAF), NSF ECCS1807216 & ECCS2032649 (BB).

AUTHOR CONTRIBUTIONS

Conceptualization, Methodology, Software, Validation, Writing—Original Draft, Writing—Review & Editing, Visualization, Formal Analysis, N.A.F., S.M., L.K., B.B., S.P., and P.O.K.; Investigation, Data Curation, N.A.F., S.M., L.K.; Information-theoretic Analysis, L.K.; Granger Causality Analysis, S.M.; Funding Acquisition, N.A.F., B.B., S.P., and P.O.K.; Resources, Supervision, Project Administration, B.B., S.P., and P.O.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURES

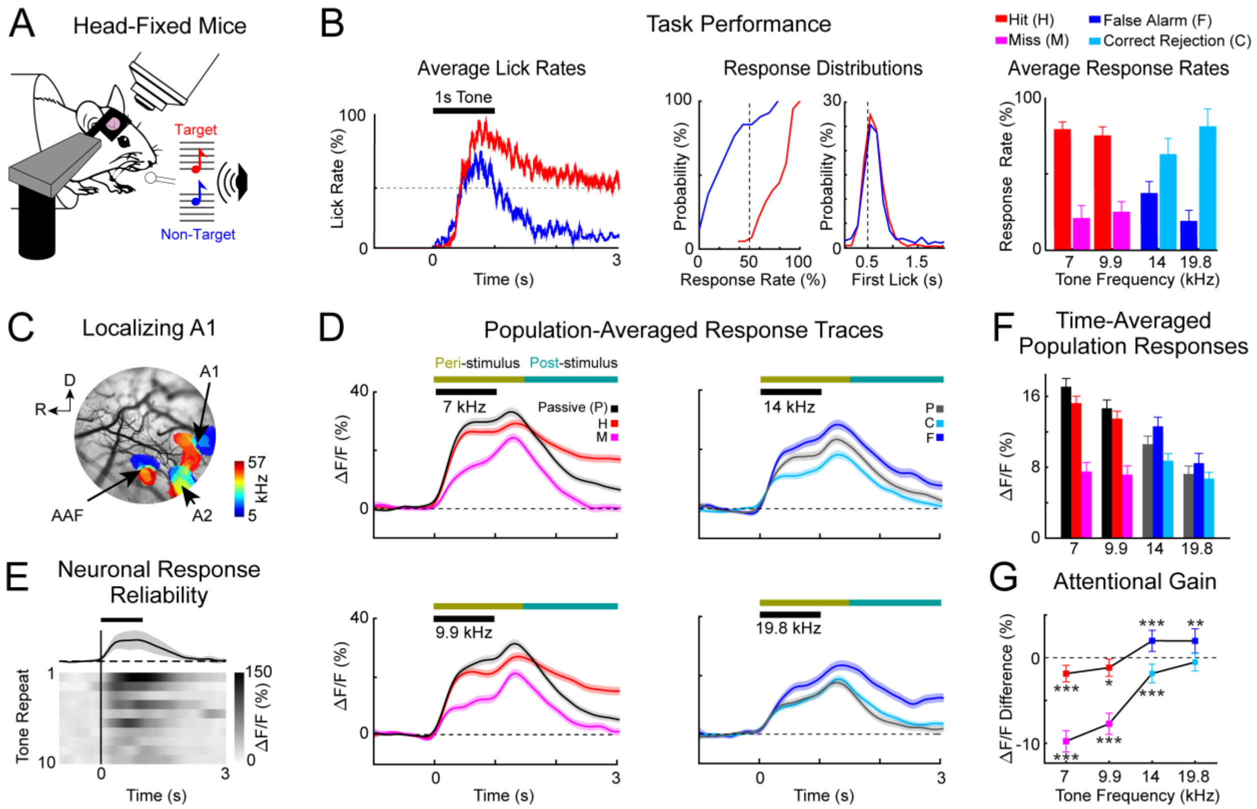


Figure 1. 2-photon imaging in awake-behaving mice shows neural responses modulated by behavioral choice. **A.** Head-fixed mice were trained to discriminate low-frequency target tones (red) vs high-frequency non-target tones (blue). **B.** Average lick rates within a trial during task performance (left panel). The horizontal black bar shows the tone presentation. The red trace (respectively, blue trace) shows the lick rate for hits (H) (respectively, false alarms (F)). The dotted line illustrates chance performance, where licking is not timed to tone presentation, but rather it is evenly distributed across a given trial. Cumulative distribution functions across experiments for hit (red) and false alarm (blue) response rates and latencies (middle panels). Average behavioral-choice rates, i.e., hit (red), miss (pink), false alarm (blue), and correct rejection (cyan), for each presented tone (right panel). Error bars show two standard errors of the mean (SEMs; n=34 experiments). **C.** Primary auditory cortex (A1) was localized within a craniotomy by using wide-field imaging to visualize tonotopy in auditory cortex. **D.** Average neuronal population response

traces in A1 layer 2/3 (L2/3) (N=2792 neurons) color-coded for behavioral choice as in panel B. Each trace shows the response to the indicated tone. Shading shows 2 SEMs. The horizontal colored bars show the peri- and post-stimulus windows, respectively, used for later analyses. **E.** Neurons in A1 L2/3 responded transiently, with jittered amplitude and timing in response to repeated identical tones. **F.** Neuronal response amplitude varied with both task performance and tone frequency. **G.** Attentional gain was defined as the difference between neural responses during behavioral vs passive trials for the same tone.

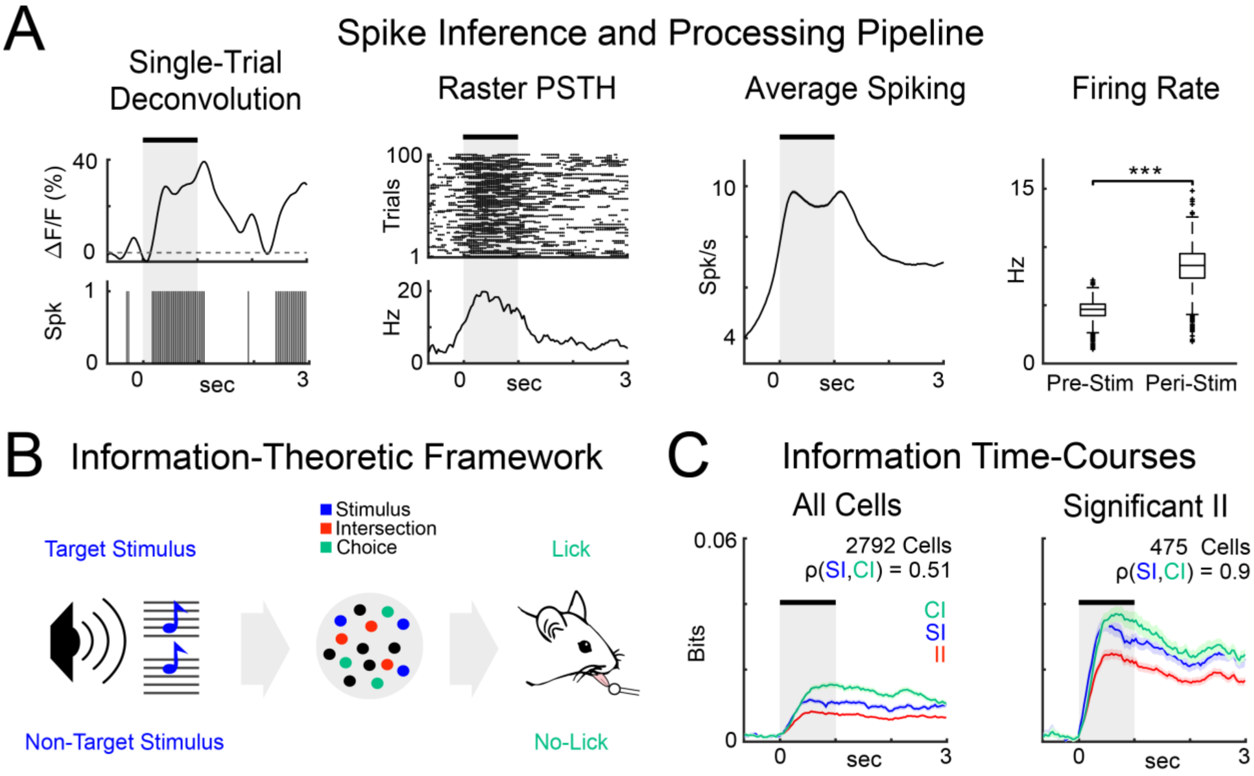


Figure 2. Processing pipeline and information-theoretic framework. **A.** Examples of deconvolution of the ΔF/F response traces (first panel); trial-by-trial spiking activity and peristimulus time histogram for a single neuron (second panel); average firing rate across neurons is higher in the post-stimulus interval than pre-stimulus (rightmost panels; p<0.001, Wilcoxon rank sum test). **B.** Stimulus encoding and behavioral read out during auditory task performance. Blue, green and red circles respectively represent neurons with stimulus information (*SI*) only, choice information

(*CI*) only, and intersection information (*II*). *II* accounts for the part of sensory and choice information used to perform the task. **C.** Time-courses of information types (*SI, CI,* and *II*) in different groups of neurons. Solid lines represent the mean and shaded areas represent SEM across all neurons in each group.

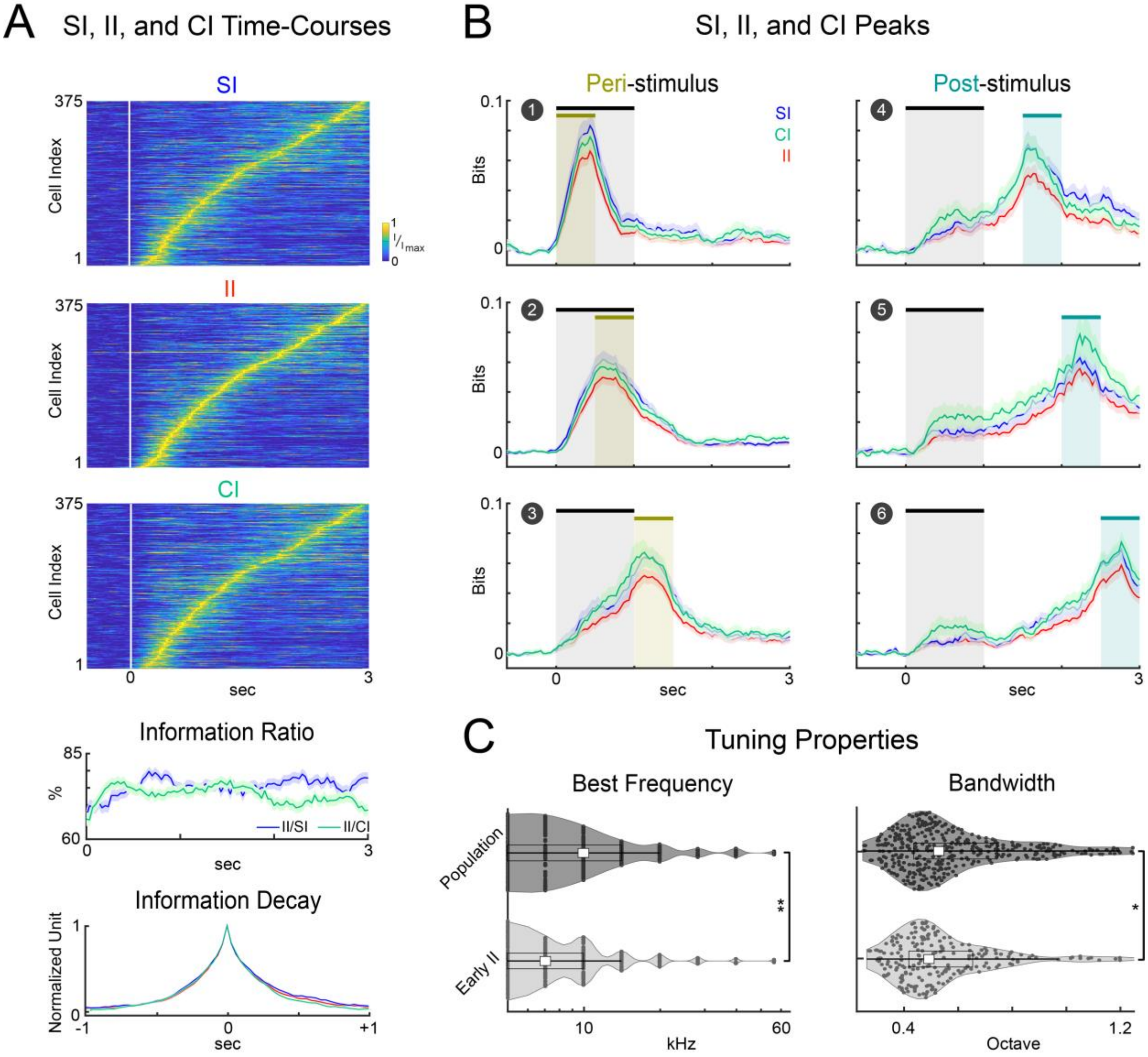


Figure 3. A1 L2/3 neurons transiently carried stimulus (*SI*), choice (*CI*) and intersection information (*II*). **A.** Information time-courses were normalized to the peak of each neuron's information and sorted by peak time of *II*. Information ratio was first computed for each neuron

and then averaged across neurons. Transiency of *SI, CI,* and *II* shown by the peak-aligned information decay within ± 1 s from the peak (bottom panel). Error bars show one SEM. **B.** Time-course of *SI, CI*, and *II* averaged over neurons. We quantified the *SI, CI,* and *II* in six separate stages of the behavioral task, which account for the *peri-stimulus* (0-1.5 s) and the *post-stimulus* intervals (1.5-3 s) shown by the shaded regions. Error bars show one SEM. **C**. Violin plots of the estimated best frequency (BF; left) and tuning bandwidth (BW; right) of neurons with early *II* vs overall population. Early *II* neurons had significantly lower BFs ($p<0.01$, Wilcoxon rank sum test) and narrower BWs ($p<0.05$, Wilcoxon rank sum test) compared to the overall population.

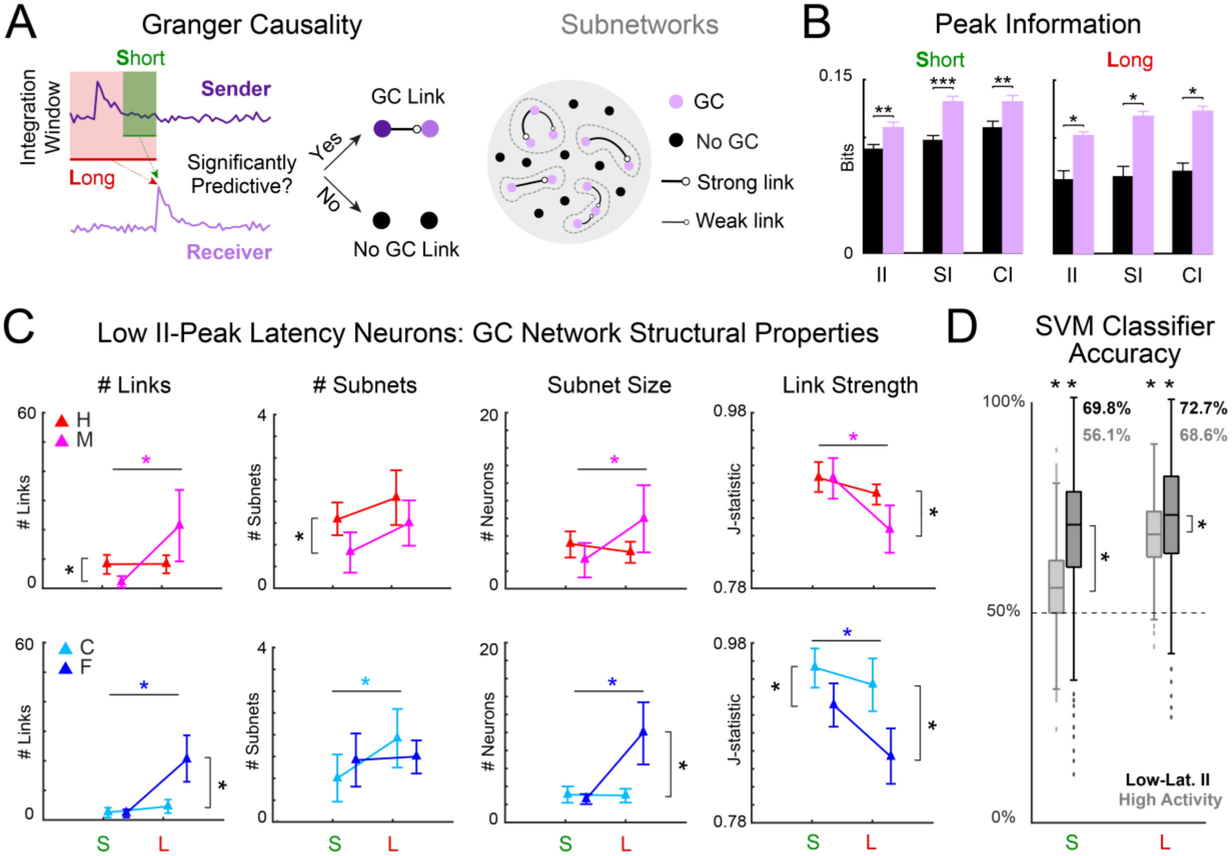


Figure 4. Behavioral choice was encoded in the network structure of low *II*-peak latency neurons. **A.** Functional networks of short (S)- and long (L)- timescale interactions amongst low *II*-peak latency neurons were estimated using Granger Causality (GC) analysis for each behavioral

choice: Hit (H), Miss (M), Correct Rejection (C), and False Alarm (F). Disjoint sets of interlinked neurons constituted subnetworks (dashed grey boundaries). **B.** GC-linked neurons, for both S and L timescales, had more information than GC-unlinked neurons (*$p<0.05$; **$p<0.01$; ***$p<0.001$). **C.** Four GC network statistics were analyzed: number of links, number of subnetworks, size of subnetworks, and statistical strength of links. Error bars show 2 SEMs. Statistically significant differences, indicated by asterisks, were identified by Wilcoxon's signed rank test ($p<0.05$). See also Supp. Table S1. **D.** Network statistics were used to train a support vector machine (SVM) to classify behavioral responses into correct or incorrect decisions. Across timescale and selection of neurons, decisions were predicted significantly better than chance ($p<0.001$). S-timescale network structure of low *II*-peak latency neurons was better decoded than highly responsive neurons ($p<0.001$). L-timescale network structures had high decoding accuracy, but low *II*-peak latency networks were better decoded than highly responsive neurons ($p<0.001$). Two-sample t-tests ($p<0.05$) were used to compare distributions and a one-sample t-test ($p<0.05$) to compare with chance performance.

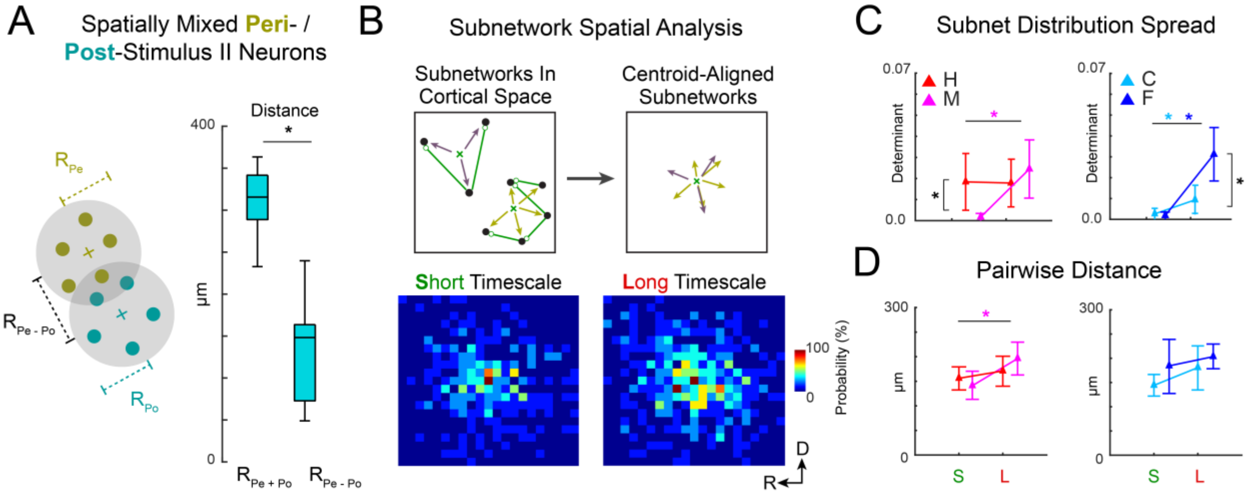


Figure 5. Subnetwork dispersion varied less by timescale during correct behavioral choices. **A.** Neurons with peri- (*Pe*) and post-stimulus (*Po*) *II* peaks were spatially intermingled. The sum of average distances of *Pe* neurons to their centroid (R_{Pe}) and of *Po* neurons to theirs (R_{Po}), denoted

as $R_{Pe + Po}$, was smaller than the distance between centroids ($R_{Pe - Po}$) ($p<0.001$, two-sample t-test). **B.** Subnetwork spatial distributions. Low *II*-peak latency neurons (black) that are linked (green) in groups isolated from others constitute subnetworks (top left). Relative locations of subnetworked neurons were aggregated over all subnetworks (top right). The distributions of relative locations are shown as 2D histograms (25 µm x 25 µm bins) for S- and L-timescales (bottom left and right). **C.** Determinant of spatial distribution covariance matrix. L-timescale C, M, and F subnetworks were more spatially dispersed than S-timescale subnetworks (M: $p<0.001$; F: $p=0.002$; C: $p=0.014$). For S-timescales, H vs M subnetworks were more dispersed ($p=0.002$), as were F vs C subnetworks for L-timescales ($p=0.003$) **D.** Pairwise distances between linked neurons remained similar for S- vs L-timescales, except for M trials ($p=0.047$). Panels **C** and **D** show mean ± 2 SEM. Asterisks indicate statistically significant differences based on Wilcoxon's signed rank test ($p<0.05$).

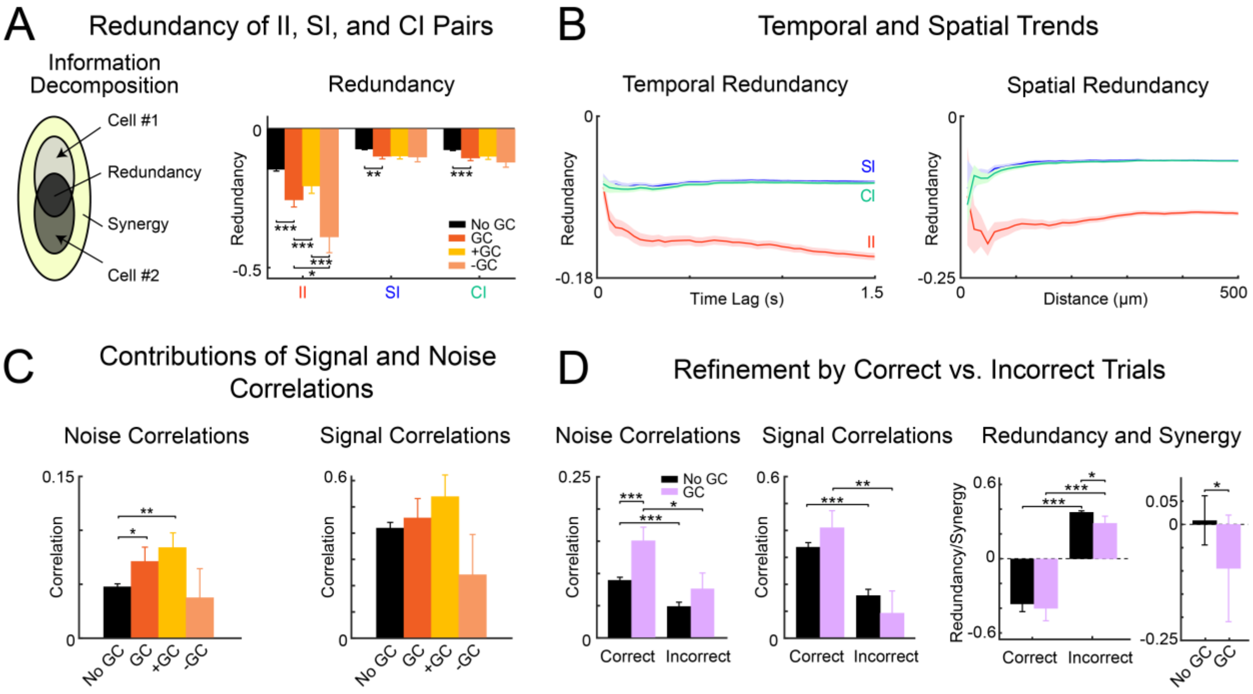


Figure 6. Redundancy and correlations increase during correct behavioral choice. **A.** Left panel: Decomposition of joint information of pairs of neurons into synergistic, cell-unique and redundant

components. Right panel: Normalized time-lagged redundancy computed for GC-linked neurons (red), either positive (orange) or negative (salmon), and GC-unlinked pairs of neurons (black). GC-linked neurons carried more redundant information than GC-unlinked neurons (*II, SI, CI*). Pairs of neurons connected with negative GC-links carried more redundant information related to *II*. **B.** Normalized redundancy across time-lagged neuronal activity (left panel), and vs the Euclidean distance (right panel) between pairs of both GC-linked and GC-unlinked neurons. **C.** Pairwise time-lagged signal and noise correlations between pairs of neurons at the peak of intersection information. Noise correlations were higher in GC-linked than GC-unlinked neurons, while signal correlations are distributed similarly. **D.** Noise and signal correlations in correct vs incorrect trials (2 leftmost panels); normalized time-lagged redundancy in correct vs incorrect trials (center-right panel); difference between the redundancy in correct vs incorrect trials for GC-linked and GC-unlinked neurons (rightmost panel). Statistical comparisons were made with a two-sample t-test (*p<0.05; **p<0.01; ***p<0.001).

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Patrick O. Kanold (pkanold@jhu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Imaging data have been deposited in the Digital Repository at the University of Maryland and are publicly available as of the date of publication. DOIs are listed in the key resources table.

- All original code has been deposited at Zenodo and the Digital Repository at the University of Maryland and is publicly available as of the date of publication. DOIs are listed in the key resources table.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All procedures were approved by the University of Maryland Institutional Animal Care and Use Committee. We used N=9 mice (3 female, 6 male) F1 offspring of CBA/CaJ strain (The Jackson Laboratory; stock #000654) crossed with transgenic C57BL/6J-Tg(thy1-GCaMP6s)GP4.3Dkim/J mice (Dana et al., 2014)(The Jackson Laboratory; stock #024275) (CBAxThy1), 8–24 weeks old, in 34 total experiments. We used the F1 generation of the crossed mice because they have good hearing into adulthood (Frisina et al., 2011). Each mouse was tested once per day over multiple days. The mice were trained to perform the task before collecting 2P data during task performance. Mice were housed under a reversed 12 h-light/12 h-dark light cycle and trained during the dark cycle.

METHOD DETAILS

Auditory task

We designed a pure-tone frequency discrimination task that used behavioral response-timing rules to induce well controlled behavioral responses in mice. Each mouse was first trained on a positive reinforcement tone detection task, with water used as a rewarding stimulus, as done previously (Francis et al., 2018b). We then trained the mice on the frequency discrimination task. Each trial began with 1 second of silence, followed by a 55 dB SPL amplitude modulated (8 Hz) tone presented for 1 s. The target tone frequencies were 7 and 9.9 kHz. The non-target frequencies were 14 and 19.8 kHz. The tone frequency was randomized across trials. The tone was followed by 2 s of silence, and a random 5-9 s inter-trial interval (ITI). The tone was presented during every trial of task-performance, and the mice were trained to lick a waterspout after the onset of a target tone and to avoid licking the waterspout after a non-target tone. Each trial's behavioral response was categorized as a hit (licking after target onset), miss (no licking after a target), false alarm (licking after non-target onset), or correct rejection (no licking in response to a non-target). Incorrect behavioral responses were punished with an 8 s time-out added to the ITI. The mice were trained to delay behavioral responses until 0.5 s after the onset of a target tone in order to be rewarded with a water droplet. While licking has been shown to impact activity in auditory cortex (Nelson and Mooney, 2016), we have previously shown that licking in the absence of perceptual decision making does not drive neural activity in A1 L2/3 during a go/no-go task (Francis et al., 2018b). Here, our use of a behavioral delay was primarily for improving task performance, since behavioral delays in a go/no-go task design reduces impulsive licking. Mice were trained on the task until hit rates were consistently above 70%, and then imaged during behavior. Mouse health was monitored daily by a skin turgor test and checking that body weight remained above 80% of the initial off-study weight.

Imaging

Chronic window implantation, widefield imaging, and 2-photon (2P) imaging, were performed as previously (Francis et al., 2018b). In brief, a chronic imaging window was implanted over a 3 mm craniotomy over auditory cortex. For widefield imaging, neuronal activity was quantified by comparing fluorescence during the stimulus versus the silent pre-stimulus baseline, resulting in a response amplitude ($\Delta F/F$). After visualizing wide-field tonotopic maps, a site was selected for 2P imaging in primary auditory cortex (A1) for each mouse. For each 2P imaging site, we determined the frequency selectivity (best frequency [BF]) of individual neurons during passive trials, i.e., trials when the mouse sat quiescently hearing tones without doing an auditory task. BFs were determined from neuronal responses to 55 dB SPL pure tones ranging from 4 - 56.6 kHz. We used a scanning microscope (Bergamo II series, B248, Thorlabs) coupled to a pulsed femtosecond Ti:Sapphire 2-photon laser with dispersion compensation (Vision S, Coherent). The microscope was controlled by ThorImageLS software. The laser was tuned to $\lambda = 940$ nm. The field of view was 370 x 370 µm. Imaging frames of 512×512 pixels (pixel size 0.72 µm) were acquired at 30 Hz by bidirectional scanning of an 8 KHz resonant scanner.

A different set of neurons was imaged for each experiment. Using an average field of view from each experiment, the somatic centers of putative neurons were manually localized and stored. A ring-like region of interest (ROI) was cropped around the cell center using the method described in Chen et al. (2013). Overlapping ROI pixels (due to closely juxtaposed neurons) were excluded from analysis. For each labeled neuron, a raw fluorescence signal over time was extracted from somatic ROIs. Pixels within the ROI were averaged to create individual neuron fluorescence traces, $F_C(t)$, for each trial of the experiment. Neuropil fluorescence was estimated for each cellular ROI using an additional ring-shaped ROI, which began 3 pixels from the somatic ROI. Pixels from the new ROI were averaged to obtain neuropil fluorescence traces, $F_N(t)$, for the same time-period as the individual neuron fluorescence traces. Pixels from regions with overlapping neuropil and cellular ROIs were removed from neuropil ROIs. Neuropil-corrected cellular fluorescence was calculated as $\hat{F}_C(t) = F_C(t) - 0.7F_N(t)$. Only cells with positive values obtained from averaging $\hat{F}_C(t)$ across time were kept for analysis, since negative values may indicate the dominance of neuropil contamination. ΔF/F was calculated from $\hat{F}_C(t)$, for each neuron, by finding the average F taken from the silent baseline period, subtracting that value from subsequent time-points, then dividing all time-points by the baseline F. All images were processed using Matlab (The Mathworks) using our prior methods (Francis et al., 2018b).

Computation of stimulus and choice information

We first deconvolved the single-trial fluorescence traces into spike rates, obtained with a sliding window approach across the entire duration of a trial (Fig. 2A). We inferred the relative spiking activity from the fluorescence traces with a first-order autoregressive model using the CaImAn algorithm (Giovannucci et al., 2019, Vogelstein et al., 2010) and binarized the deconvolved traces into 0, when there was no activity at all, and 1, if the spiking activity was above 0 (Fig. 2A). To validate our deconvolution, we computed the averaged spiking activity time-course of the whole population. In accordance with previous studies (Forli et al., 2018, Petrus et al., 2014), we found the firing rates in the pre-/peri-stimulus intervals were 4.4 ± 0.7 and 8.2 ± 1.6 Hz (mean $\pm$ s.d. across cells) respectively (Fig. 2A). We computed mutual information carried by neurons at a given time either about stimulus category S (low vs high frequency tones), and about the behavioral choices C (lick vs. no-lick), defined as follows (Quian Quiroga and Panzeri, 2009, Cover and Thomas, 1991):

$$I(X;R_t) = \sum_{x,r_t} p(x,r_t)\log_2\left[\frac{p(x,r_t)}{p(x)p(r_t)}\right] \qquad (1)$$

where $X = S, C$ denotes the set of task variables, either stimuli S or choices C. R_t is the set of responses of the neuron measured at a given time t. $p(x,r_t)p(r_t)$ denotes the joint probability of observing in a given trial a value x for the stimulus or choice variable and a value $r_{n,t}$ for the activity of the neuron at time t. $p(x) = \sum_r p(x,r_t)$, and $p(r_t) = \sum_x p(x,r_t)$, are the marginal probabilities.

To compute the time dependent spike rates r_t, we first inferred the relative spiking activity from the fluorescence traces with the CaImAn first-order autoregressive model (Giovannucci et al., 2019, Vogelstein et al., 2010). We then averaged the spiking activities with a sliding window of 10 imaging frames, in time-steps of 1 imaging frame, and the resulting activity was binarized as 0/1 (occurrence or not of at least 1 spike in each window). We then computed information in these spike rates from Eq (1) with the Information Breakdown Toolbox (Magri et al., 2009). We then subtracted for each neuron the average information computed in the pre-stimulus interval. This procedure insures the removal of the limited sampling bias, as well as the removal of possible contributions of pre-stimulus choice signals (Niwa et al., 2013) reflecting e.g. stimulus-unrelated internal bias.

Computation of intersection information

We computed intersection information $II(S,R,C)$, following exactly the procedure published in (Pica et al., 2017), to which we refer for full details. $II(S,R,C)$ quantifies the part of sensory information that is is used to inform behavioral readout and is bounded by both $I(S;R)$ and choice $I(S;C)$ (Pica et al., 2017). As for stimulus and choice information, we subtracted the average intersection information evaluated in the pre-stimulus interval to remove limited sampling biases.

Computation of information significance

To select individual neurons with significant information of each type, we used a non-parametric permutation test (creating a null hypothesis distribution of information values obtained randomly shuffling across trials the stimulus-response or choice-response associations), and we set a threshold of p<0.1. Note that we constructed the null hypothesis distribution selecting for each random permutation the maximum information over all time windows of the permuted values. The so obtained p values are already corrected for multiple comparisons across time bins. Because having significant *II* also requires having significant *CI* and *SI*, when selecting neurons with significant *II* we required that those neurons have also significant *SI*, *CI* and *II* at p<0.1. Given that these three tests are not independent, we could not evaluate the expected number of falsely labeled *II* neurons simply by taking the product of the three p-value thresholds. To empirically estimate the rate of falsely labeled *II* neurons, we repeated our selection procedure using surrogate data in which for each neuron we randomly permuted the trials at the outset. The fraction of surrogate neurons that exhibit significance at p<0.1 simultaneously for *CI*, *SI* and *II* provides an empirical estimate of the false positive rate in our selection procedure. We found that only 1% (31 out of 2792 neurons) of the surrogate neurons were classified as significant. Thus, our set of II-selected neurons contains 1% of falsely labeled *II* neurons, which corresponds effectively to a p<0.01 selection p-value.

We further checked how our results generalize when requiring at same time significant *SI, CI* and *II* with a more stringent threshold at p<0.05. In this case, we could only select 7 (rather than 12) experimental sessions with at least 20 *II* neurons for GC analysis. We repeated our analysis for *II* neurons chosen with threshold p<0.05 (See Supp. Fig. S3E-F and Supp. Table S5), finding the same qualitative trends of Fig. 4 but with fewer significant differences due to fewer number of sessions (N=7 in Supp. Fig. S3E-F vs. N=12 in Fig. 3). Notably, we were able to decode correct vs. incorrect choice more accurately (Supp. Fig. S3F) with this even more stringent threshold.

Granger Causality Analysis

Granger causality (GC) analysis evaluates the predictive influence of the past activity of one neural process on present activity of another. GC analysis was performed similarly as in our previous work (Francis et al., 2018b) by fitting sparse vector autoregressive (VAR) models to the ensemble neural responses (ΔF/F), calculating an unbiased GC measure for each potential link, and characterizing the GC link strengths using Youden's J-statistics following false discovery rate control at a rate of 0.001. We highlight here three key differences from previous analysis regarding model estimation, modelling history-dependency, and neuron selection, and refer the reader to (Francis et al., 2018b) for a recapitulation of the remaining details.

In order to estimate GC network connectivity amongst larger networks, the maximum likelihood problem in (Francis et al., 2018b) is solved, employing the Orthogonal Matching Pursuit (OMP) algorithm (Zhang, 2011, Cai and Wang, 2011) to fit sparse VAR models rather than ℓ_1-regularisation. OMP enables the sparsity of the estimated parameter vector—i.e. the number of non-zero parameters—to be controlled, thus mitigating model overfitting more robustly. The sparsity level of each VAR model is obtained by cross-validation. The set of non-zero parameters, called the model support set, is iteratively selected: at each iteration, a new parameter with the greatest contribution to the residual estimation error is added to the support and maximum likelihood estimation is performed over the updated support set.

The neural responses of a set of C neurons, indexed by $c = 1, \dots, C$, are denoted by $\left\{y_{r,n}^{(c)}\right\}_{r=1:R,n=1:N}^{c=1:C}$, where $n = 1, \dots, N$ and $r = 1, \dots, R$ index time bins and trial repetitions, respectively. The covariates of the VAR model of each neural response incorporate the self- and cross-histories of activity over an integration window of L samples within which neuronal interactions are assumed to occur. The integration window is subdivided into M non-overlapping windows of lengths $\{W_m\}_{m=1:M}$. The average activity of neuron (c) in the m-th window lag with respect to time bin n and trial r is given by

$$h_{r,n,m}^{(c)} = \frac{1}{W_m} \sum\nolimits_{k=n-1-b_m}^{n-1-b_{m-1}} y_{r,k}^{(c)}, \qquad (2)$$

where $b_m = \sum_{l=1}^{m} W_m$ and $b_0 = 0$. The collection of history covariates $\left\{h_{r,n,m}^{(c)}\right\}_{m=1:M}^{c=1:C}$ comprises the regressors of $y_{r,n}^{(c)}$. Note that the conditional independence of responses given the collection of history covariates allows to estimate the VAR parameters at the single-trial level by maximizing the joint likelihood of the within-trial responses via OMP (Francis et al., 2018b).

Interactions between neurons over short timescales (S) are modelled with an integration window of $L = 7$ lags with $M = 3$ subdivisions with window lengths $\{W_m\}_{m=1:M} = \{2^{m-1}\}_{m=1:M}$ lags. Long timescale (L) interactions are modelled by instead using a cross-history integration window of length $L = 31$ lags with $M = 5$ subdivisions whose window lengths are similarly defined. S and L timescale interactions thus respectively correspond to 233 ms and 1033 ms windows of effective history. It is clear from the parameterization that the S and L interactions are modelled in a nested fashion. We validate this approach by simulating a 10 neuron network consisting of both S- and L-timescale links (see Supp. Fig. S6). Employing the L integration window for GC analysis, we are able to correctly identify all L- and S-timescale interactions; however, using the S integration window, while the S-timescale links are correctly identified, the L-timescale links are expectedly discarded, thus corroborating the sensitivity and specificity of our proposed inference framework.

Twenty neurons were analyzed from each 2P experiment. Analyzing a subset of fixed size avoids intersession variations in the number of recorded neurons that could affect analyses. The total number of model parameters, $M \cdot C$, needs to be much smaller than the total number of samples, $R \cdot N$, for reliable model estimation. We use at most $M = 5$ subintervals and per trial used the $N = 105$ time samples of the response after stimulus onset; we calculated $C = 20$ to be the maximum number of neurons that satisfies this condition, conservatively assuming at minimum $R = 10$ trials per session of each behavioral choice category. In our main results, 20 *II* neurons with the lowest *II*-peak latency in each experiment (N=12) in which at least as many *II* neurons were identified. For consistency, 20 exclusively *CI* neurons were similarly selected in the analyses presented in Supp. Fig. S3A-B. Highly active neurons in each 2P experiment (N=34) were selected per behavioral choice category. The neural response of the c^{th} neuron at the n^{th} time index of the r^{th} repeated trial of a behavioral category, $y_{r,n}^{(c)}$, is normalized $\tilde{y}_{r,n}^{(c)} = \frac{y_{r,n}^{(c)}}{\sqrt{\sum_n (y_{r,n}^{(c)})^2}}$.

The 20 neurons with the smallest trial-averaged variances of the normalized responses, $\overline{\sigma}^{(c)^2} = \frac{1}{R}\sum_r Var(y_{r,n}^{(c)})$, were selected.

Decoding behavioral choice from network structure

To test if network structures encode behavioral choice, we trained classifiers on four GC network statistics — number of links, number of subnetworks, size of subnetworks, and statistical strength of links — to distinguish correct (Hit or Correct Rejection) and incorrect (Miss or False Alarm) decisions. Feature vectors consisting of these statistics were compiled for each behavioral choice network from the VAR parameters estimated at the single-trial level. We then trained a linear support vector machine (SVM) at the single-trial level to predict behavioral choice using a randomly selected 75% of the feature vectors, with the remaining 25% used to evaluate prediction accuracy. This procedure was repeated 2000 times, each with a new randomized partition of feature vectors, to characterize the distribution of average classification accuracy.

Spatial Distribution of GC subnetworks

To investigate the spatial scales over which functionally linked neurons interact, we leveraged the spatial location of individual neurons available in 2P imaging to analyze how subnetworks were distributed across the imaged cortical area. To this end, the locations of subnetworked neurons relative to their centroid were obtained as follows. For a subnetwork of R neurons with positions $\{(x_i, y_i)\}_{i=1}^{R}$, we compute their locations relative to the subnetwork centroid, $\{(x_i - \overline{x}, y_i - \overline{y})\}_{i=1}^{R}$, where $\overline{x} = \frac{1}{R}\sum_{i=1}^{r} x_i$ and $\overline{y} = \frac{1}{R}\sum_{i=1}^{r} y_i$. The relative locations are compiled over all subnetworks to yield an empirical distribution. The covariance matrix of the distribution describes the spatial spread of subnetworks. Its determinant – which accounts for both the covariance between x and y as well as their respective variances – is used as a comparative statistic to quantify differences in the spatial dispersion of subnetworks across conditions.

Computation of information redundancy

We used a normalized redundancy index defined as the information carried jointly by two neurons minus the sum of the information that each carried independently, normalized with respect to the total information carried by the two neurons jointly (Pola et al., 2003, Schneidman et al., 2003):

$$Redundancy = \frac{I\left(X; R_{1,t_1}, R_{2,t_2}\right) - I\left(S; R_{1,t_1}\right) - I\left(S; R_{2,t_2}\right)}{I\left(X; R_{1,t_1}, R_{2,t_2}\right)} \qquad (3)$$

For each of the two neurons, we selected activity at the time t_1,t_2 of their peak information. The single neuron information was computed as in section “Definition of stimulus, choice and intersection information for single neurons”. The joint time-lagged stimulus and choice information was computed as follows:

$$I\left(X; R_{1,t_1}, R_{2,t_2}\right) = \sum_{x} \sum_{r_{1,t_1}, r_{2,t_2}} p\left(x, r_{1,t_1}, r_{2,t_2}\right) \log_2 \left[\frac{p\left(x, r_{1,t_1}, r_{2,t_2}\right)}{p(x) p\left(r_{1,t_1}, r_{2,t_2}\right)}\right] \qquad (4)$$

Notations are as in Eq. (1) , with now $p\left(X, r_{1,t_1}, r_{2,t_2}\right)$ denoting the probability of observing in a given trial a value x of the behavioral variable (stimulus category or choice) and a joint response r_{1,t_1}, r_{2,t_2} of the two neurons at times t_1,t_2 respectively. Intersection information was computed with the methods detailed in (Pica et al., 2017), using the joint response r_{1,t_1}, r_{2,t_2} as the neural response variable.

Computation of signal and noise correlations

We computed noise correlations as across-trials Pearson correlations of the activity of pairs of neurons at fixed stimulus (then averaged over stimuli), whereas signal correlations were computed as Pearson correlations across stimuli of the trial-averaged responses to each of the two stimuli (high vs low frequency tones). Given the imbalance in the number of correct and incorrect trials, we equalized the sample sizes by randomly subsampling the correct trials, to avoid systematic errors in the comparisons between correct and incorrect trials (see Fig. 6D).

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless noted otherwise, statistical comparisons were performed using a bootstrap t-test with 10000 iterations or a Kolmogorov–Smirnov test (KS-test), for both one- and paired-sample tests. Kruskal-Wallis tests were used when there were >2 groups being compared. We used a Bonferroni correction for multiple comparisons. All mean values are reported with 2 standard errors of the mean, unless noted differently.

REFERENCES

ALONSO, J.-M., USREY, W. M. & REID, R. C. 1996. Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature,* 383**,** 815-819.

ATENCIO, C. A. & SCHREINER, C. E. 2010. Columnar connectivity and laminar processing in cat primary auditory cortex. *PLoS One,* 5**,** e9521.

ATTNEAVE, F. 1954. Some informational aspects of visual perception. *Psychological review,* 61**,** 183-193.

ATZORI, M., LEI, S., EVANS, D. I., KANOLD, P. O., PHILLIPS-TANSEY, E., MCINTYRE, O. & MCBAIN, C. J. 2001. Differential synaptic processing separates stationary from transient inputs to the auditory cortex. *Nat Neurosci,* 4**,** 1230-7.

BAGUR, S., AVERSENG, M., ELGUEDA, D., DAVID, S., FRITZ, J., YIN, P., SHAMMA, S., BOUBENEC, Y. & OSTOJIC, S. 2018. Go/No-Go task engagement enhances population representation of target stimuli in primary auditory cortex. *Nat Commun,* 9**,** 2529.

BANDYOPADHYAY, S., SHAMMA, S. A. & KANOLD, P. O. 2010. Dichotomy of functional organization in the mouse auditory cortex. *Nat Neurosci,* 13**,** 361-8.

BARLOW, H. B. 1961. Possible principles underlying the transformations of sensory messages. *Sensory communication***,** 217-234.

BERTSCHINGER, N., RAUH, J., OLBRICH, E., JOST, J. & AY, N. 2014. Quantifying Unique Information. *Entropy,* 16**,** 2161-2183.

BRESSLER, S. L. & SETH, A. K. 2011. Wiener-Granger causality: a well established methodology. *Neuroimage,* 58**,** 323-9.

BROSCH, M., SELEZNEVA, E. & SCHEICH, H. 2011. Representation of reward feedback in primate auditory cortex. *Front Syst Neurosci,* 5**,** 5.

CAI, T. T. & WANG, L. 2011. Orthogonal Matching Pursuit for Sparse Signal Recovery With Noise. *IEEE Transactions on Information Theory,* 57**,** 4680-4688.

CARCEA, I., INSANALLY, M. N. & FROEMKE, R. C. 2017. Dynamics of auditory cortical activity during behavioural engagement and auditory perception. *Nat Commun,* 8**,** 14412.

CHECHIK, G., ANDERSON, M. J., BAR-YOSEF, O., YOUNG, E. D., TISHBY, N. & NELKEN, I. 2006. Reduction of information redundancy in the ascending auditory pathway. *Neuron,* 51**,** 359-68.

CHEN, I. W., HELMCHEN, F. & LUTCKE, H. 2015. Specific Early and Late Oddball-Evoked Responses in Excitatory and Inhibitory Neurons of Mouse Auditory Cortex. *J Neurosci,* 35**,** 12560-73.

CHEN, T. W., WARDILL, T. J., SUN, Y., PULVER, S. R., RENNINGER, S. L., BAOHAN, A., SCHREITER, E. R., KERR, R. A., ORGER, M. B., JAYARAMAN, V., LOOGER, L. L., SVOBODA, K. & KIM, D. S. 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature,* 499**,** 295-300.

CHICHARRO, D., PANZERI, S. & HAEFNER, R. 2021. Stimulus-dependent relationships between behavioral choice and sensory neural responses. *Elife,* 10.

CHRISTISON-LAGAY, K. L. & COHEN, Y. E. 2018. The Contribution of Primary Auditory Cortex to Auditory Categorization in Behaving Monkeys. *Front Neurosci,* 12**,** 601.

COVER, T. M. & THOMAS, J. A. 1991. *Elements of information theory,* New York, Wiley.

DANA, H., CHEN, T. W., HU, A., SHIELDS, B. C., GUO, C., LOOGER, L. L., KIM, D. S. & SVOBODA, K. 2014. Thy1-GCaMP6 transgenic mice for neuronal population imaging in vivo. *PLoS One,* 9**,** e108697.

DAVID, S. V., FRITZ, J. B. & SHAMMA, S. A. 2012. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 2144-2149.

FORLI, A., VECCHIA, D., BININI, N., SUCCOL, F., BOVETTI, S., MORETTI, C., NESPOLI, F., MAHN, M., BAKER, C., BOLTON, M., YIZHAR, O. & FELLIN, T. 2018. Two-Photon Bidirectional Control and Imaging of Neuronal Excitability with High Spatial Resolution In Vivo. *Cell Reports,* 22**,** 3087-3098.

FRANCIS, N. A., ELGUEDA, D., ENGLITZ, B., FRITZ, J. B. & SHAMMA, S. A. 2018a. Laminar profile of task-related plasticity in ferret primary auditory cortex. *Sci Rep,* 8**,** 16375.

FRANCIS, N. A., WINKOWSKI, D. E., SHEIKHATTAR, A., ARMENGOL, K., BABADI, B. & KANOLD, P. O. 2018b. Small Networks Encode Decision-Making in Primary Auditory Cortex. *Neuron,* 97**,** 885-897 e6.

FRISINA, R. D., SINGH, A., BAK, M., BOZORG, S., SETH, R. & ZHU, X. 2011. F1 (CBAxC57) mice show superior hearing in old age relative to their parental strains: hybrid vigor or a new animal model for "golden ears"? *Neurobiol Aging,* 32**,** 1716-24.

FRITZ, J., SHAMMA, S., ELHILALI, M. & KLEIN, D. 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci,* 6**,** 1216-23.

FRITZ, J. B., DAVID, S. V., RADTKE-SCHULLER, S., YIN, P. & SHAMMA, S. A. 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat Neurosci,* 13**,** 1011-9.

GIOVANNUCCI, A., FRIEDRICH, J., GUNN, P., KALFON, J., BROWN, B. L., KOAY, S. A., TAXIDIS, J., NAJAFI, F., GAUTHIER, J. L., ZHOU, P., KHAKH, B. S., TANK, D. W., CHKLOVSKII, D. B. & PNEVMATIKAKIS, E. A. 2019. CaImAn an open source tool for scalable calcium imaging data analysis. *Elife,* 8.

GOLD, J. I. & SHADLEN, M. N. 2001. Neural computations that underlie decisions about sensory stimuli. *Trends Cogn Sci,* 5**,** 10-16.

GUO, L., WEEMS, J. T., WALKER, W. I., LEVICHEV, A. & JARAMILLO, S. 2019. Choice-Selective Neurons in the Auditory Cortex and in Its Striatal Target Encode Reward Expectation. *J Neurosci,* 39**,** 3687-3697.

INSANALLY, M. N., CARCEA, I., FIELD, R. E., RODGERS, C. C., DEPASQUALE, B., RAJAN, K., DEWEESE, M. R., ALBANNA, B. F. & FROEMKE, R. C. 2019. Spike-timing-dependent ensemble encoding by non-classically responsive cortical neurons. *Elife,* 8.

KAMINSKI, M., DING, M., TRUCCOLO, W. A. & BRESSLER, S. L. 2001. Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance. *Biol Cybern,* 85**,** 145-57.

KANOLD, P. O., NELKEN, I. & POLLEY, D. B. 2014. Local versus global scales of organization in auditory cortex. *Trends Neurosci,* 37**,** 502-10.

KATO, H. K., GILLET, S. N. & ISAACSON, J. S. 2015. Flexible Sensory Representations in Auditory Cortex Driven by Behavioral Relevance. *Neuron,* 88.

KHOURI, L. & NELKEN, I. 2015. Detecting the unexpected. *Curr Opin Neurobiol,* 35**,** 142-7.

KIM, S., PUTRINO, D., GHOSH, S. & BROWN, E. N. 2011. A Granger causality measure for point process models of ensemble neural spiking activity. *PLoS Comput Biol,* 7**,** e1001110.

KUCHIBHOTLA, K. V., GILL, J. V., LINDSAY, G. W., PAPADOYANNIS, E. S., FIELD, R. E., STEN, T. A., MILLER, K. D. & FROEMKE, R. C. 2017. Parallel processing by cortical inhibition enables context-dependent behavior. *Nat Neurosci,* 20**,** 62-71.

LIU, J. & KANOLD, P. O. 2021. Diversity of Receptive Fields and Sideband Inhibition with Complex Thalamocortical and Intracortical Origin in L2/3 of Mouse Primary Auditory Cortex. *J Neurosci,* 41**,** 3142-3162.

LIU, J., WHITEWAY, M. R., SHEIKHATTAR, A., BUTTS, D. A., BABADI, B. & KANOLD, P. O. 2019. Parallel Processing of Sound Dynamics across Mouse Auditory Cortex via Spatially Patterned Thalamic Inputs and Distinct Areal Intracortical Circuits. *Cell Rep,* 27**,** 872-885 e7.

LIU, Y., XIN, Y. & XU, N. L. 2021. A cortical circuit mechanism for structural knowledge-based flexible sensorimotor decision-making. *Neuron,* 109**,** 2009-2024 e6.

MAGRI, C., WHITTINGSTALL, K., SINGH, V., LOGOTHETIS, N. K. & PANZERI, S. 2009. A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. *BMC Neurosci,* 10**,** 81.

MAKKEH, A., THEIS, D. & VICENTE, R. 2018. BROJA-2PID: A Robust Estimator for Bivariate Partial Information Decomposition. *Entropy,* 20.

MAOR, I., SHALEV, A. & MIZRAHI, A. 2016. Distinct Spatiotemporal Response Properties of Excitatory Versus Inhibitory Neurons in the Mouse Auditory Cortex. *Cereb Cortex,* 26**,** 4242-4252.

MCGINLEY, M. J., DAVID, S. V. & MCCORMICK, D. A. 2015. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron,* 87**,** 179-92.

MENG, X., WINKOWSKI, D. E., KAO, J. P. Y. & KANOLD, P. O. 2017. Sublaminar Subdivision of Mouse Auditory Cortex Layer 2/3 Based on Functional Translaminar Connections. *J Neurosci,* 37**,** 10200-10214.

NELSON, A. & MOONEY, R. 2016. The Basal Forebrain and Motor Cortex Provide Convergent yet Distinct Movement-Related Inputs to the Auditory Cortex. *Neuron,* 90**,** 635-48.

NIGAM, S., POJOGA, S. & DRAGOI, V. 2019. Synergistic Coding of Visual Information in Columnar Networks. *Neuron,* 104**,** 402-411 e4.

NIRENBERG, S., CARCIERI, S. M., JACOBS, A. L. & LATHAM, P. E. 2001. Retinal ganglion cells act largely as independent encoders. *Nature,* 411**,** 698-701.

NIWA, M., JOHNSON, J. S., O'CONNOR, K. N. & SUTTER, M. L. 2013. Differences between primary auditory cortex and auditory belt related to encoding and choice for AM sounds. *J Neurosci,* 33**,** 8378-95.

OVIEDO, H. V., BUREAU, I., SVOBODA, K. & ZADOR, A. M. 2010. The functional asymmetry of auditory cortex is reflected in the organization of local cortical circuits. *Nat Neurosci,* 13**,** 1413-20.

PANZERI, S., HARVEY, C. D., PIASINI, E., LATHAM, P. E. & FELLIN, T. 2017. Cracking the Neural Code for Sensory Perception by Combining Statistics, Intervention, and Behavior. *Neuron,* 93**,** 491-507.

PARRAS, G. G., CASADO-ROMAN, L., SCHROGER, E. & MALMIERCA, M. S. 2021. The posterior auditory field is the chief generator of prediction error signals in the auditory cortex. *Neuroimage,* 242**,** 118446.

PETRUS, E., ISAIAH, A., JONES, A., LI, D., WANG, H., LEE, H. & KANOLD, P. 2014. Crossmodal Induction of Thalamocortical Potentiation Leads to Enhanced Information Processing in the Auditory Cortex. *Neuron,* 81**,** 664-673.

PI, H. J., HANGYA, B., KVITSIANI, D., SANDERS, J. I., HUANG, Z. J. & KEPECS, A. 2013. Cortical interneurons that specialize in disinhibitory control. *Nature,* 503**,** 521-4.

PICA, G., PIASINI, E., SAFAAI, H., RUNYAN, C., DIAMOND, M., FELLIN, T., KAYSER, C., HARVEY, C. & PANZERI, S. 2017. Quantifying how much sensory information in a neural code is relevant for behavior. *Advances in Neural Information Processing Systems 30 (NeurIPS)***,** 3686–3696.

POLA, G., THIELE, A., HOFFMANN, K. P. & PANZERI, S. 2003. An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network,* 14**,** 35-60.

QUIAN QUIROGA, R. & PANZERI, S. 2009. Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci,* 10**,** 173-85.

QUINN, C. J., KIYAVASH, N. & COLEMAN, T. P. 2015. Directed Information Graphs. *IEEE Transactions on Information Theory,* 61**,** 6887-6909.

QUIROGA, R. & PANZERI, S. 2009. Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience,* 10**,** 173-185.

REICH, D. S., MECHLER, F. & VICTOR, J. D. 2001. Independent and redundant information in nearby cortical neurons. *Science,* 294**,** 2566-8.

RODGERS, C. C. & DEWEESE, M. R. 2014. Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task for rodents. *Neuron,* 82**,** 1157-70.

ROTHSCHILD, G., NELKEN, I. & MIZRAHI, A. 2010. Functional organization and population dynamics in the mouse primary auditory cortex. *Nat Neurosci,* 13**,** 353-60.

RUNYAN, C. A., PIASINI, E., PANZERI, S. & HARVEY, C. D. 2017. Distinct timescales of population coding across cortex. *Nature,* 548**,** 92-96.

SADAGOPAN, S. & WANG, X. 2009. Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. *J Neurosci,* 29**,** 11192-202.

SAKATA, S. & HARRIS, K. D. 2009. Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron,* 64**,** 404-18.

SALINAS, E. & SEJNOWSKI, T. J. 2001. Correlated neuronal activity and the flow of neural information. *Nature reviews neuroscience,* 2**,** 539-550.

SCHNEIDMAN, E., BIALEK, W. & BERRY, M. J. 2003. Synergy, Redundancy, and Independence in Population Codes. *The Journal of Neuroscience,* 23**,** 11539.

SCHWARTZ, Z. P. & DAVID, S. V. 2018. Focal Suppression of Distractor Sounds by Selective Attention in Auditory Cortex. *Cereb Cortex,* 28**,** 323-339.

SETH, A. K., BARRETT, A. B. & BARNETT, L. 2015. Granger causality analysis in neuroscience and neuroimaging. *J Neurosci,* 35**,** 3293-7.

SHANNON, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal,* 27**,** 379-423.

SHEIKHATTAR, A., MIRAN, S., LIU, J., FRITZ, J. B., SHAMMA, S. A., KANOLD, P. O. & BABADI, B. 2018. Extracting neuronal functional network dynamics via adaptive Granger causality analysis. *Proc Natl Acad Sci U S A,* 115**,** E3869-E3878.

STEINMETZ, N. A., ZATKA-HAAS, P., CARANDINI, M. & HARRIS, K. D. 2019. Distributed coding of choice, action and engagement across the mouse brain. *Nature,* 576**,** 266-273.

STRINGER, C., PACHITARIU, M., STEINMETZ, N., REDDY, C. B., CARANDINI, M. & HARRIS, K. D. 2019. Spontaneous behaviors drive multidimensional, brainwide activity. *Science,* 364**,** 255.

TSUNADA, J., LIU, A. S. K., GOLD, J. I. & COHEN, Y. E. 2016. Causal contribution of primate auditory cortex to auditory perceptual decision-making (vol 19, pg 135, 2015). *Nature Neuroscience,* 19**,** 642-642.

VALENTE, M., PICA, G., BONDANELLI, G., MORONI, M., RUNYAN, C. A., MORCOS, A. S., HARVEY, C. D. & PANZERI, S. 2021. Correlations enhance the behavioral readout of neural population activity in association cortex. *Nature Neuroscience,* 24**,** 975-986.

VOGELSTEIN, J. T., PACKER, A. M., MACHADO, T. A., SIPPY, T., BABADI, B., YUSTE, R. & PANINSKI, L. 2010. Fast Nonnegative Deconvolution for Spike Train Inference From Population Calcium Imaging. *Journal of Neurophysiology,* 104**,** 3691-3704.

WILLIAMS, P. L. & BEER, R. D. 2010. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*.

WINKOWSKI, D. E., BANDYOPADHYAY, S., SHAMMA, S. A. & KANOLD, P. O. 2013. Frontal cortex activation causes rapid plasticity of auditory cortical processing. *J Neurosci,* 33**,** 18134-48.

WINKOWSKI, D. E. & KANOLD, P. O. 2013. Laminar transformation of frequency organization in auditory cortex. *J Neurosci,* 33**,** 1498-508.

WINKOWSKI, D. E., NAGODE, D. A., DONALDSON, K. J., YIN, P., SHAMMA, S. A., FRITZ, J. B. & KANOLD, P. O. 2018. Orbitofrontal Cortex Neurons Respond to Sound and Activate Primary Auditory Cortex Neurons. *Cereb Cortex,* 28**,** 868-879.

YANG, Y., DEWEESE, M. R., OTAZU, G. H. & ZADOR, A. M. 2008. Millisecond-scale differences in neural activity in auditory cortex can drive decisions. *Nat Neurosci,* 11**,** 1262-3.

YIN, P., STRAIT, D. L., RADTKE-SCHULLER, S., FRITZ, J. B. & SHAMMA, S. A. 2020. Dynamics and Hierarchical Encoding of Non-compact Acoustic Categories in Auditory and Frontal Cortex. *Curr Biol,* 30**,** 1649-1663 e5.

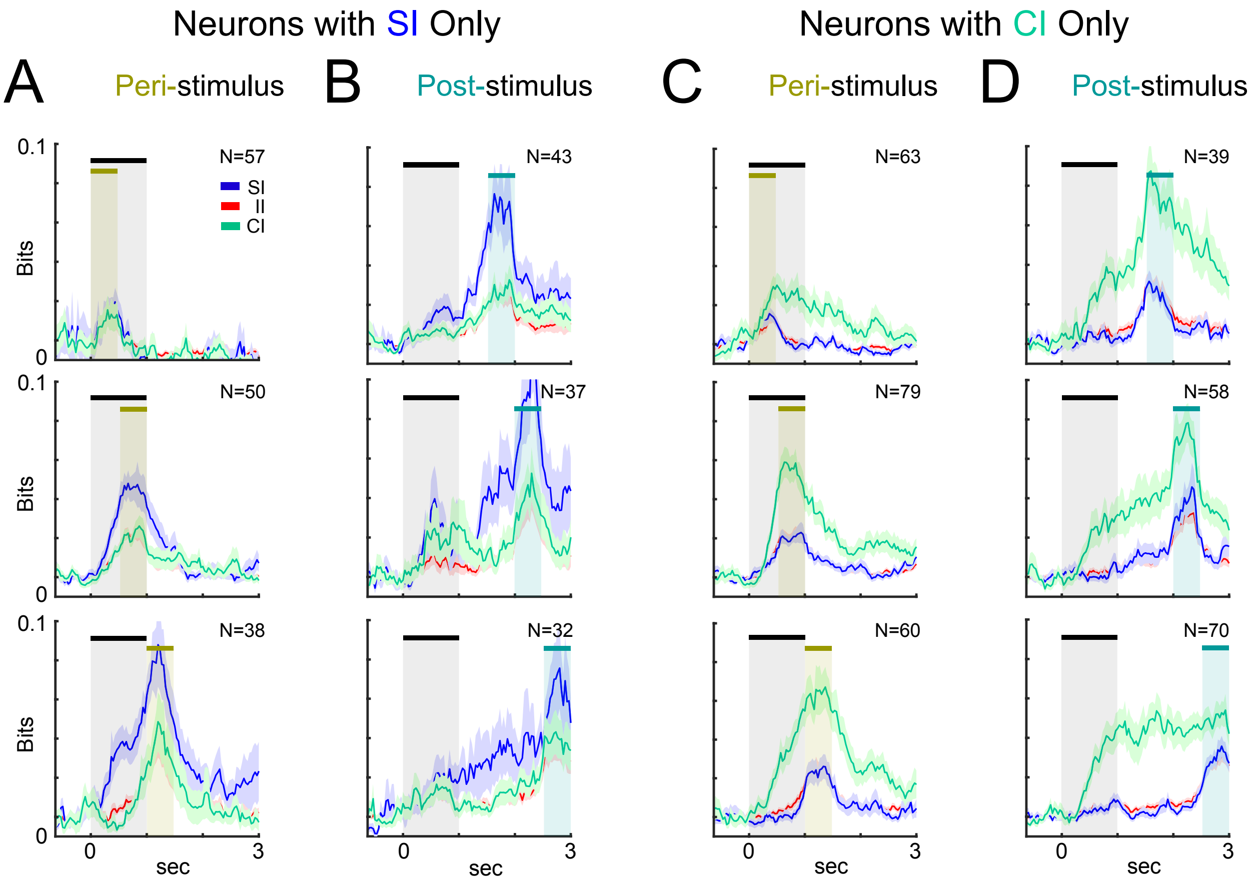
ZHANG, T. 2011. Sparse Recovery With Orthogonal Matching Pursuit Under RIP. *IEEE Transactions on Information Theory,* 57**,** 6215-6221.

Sequential Transmission of Task-Relevant Information in Cortical Neuronal Networks

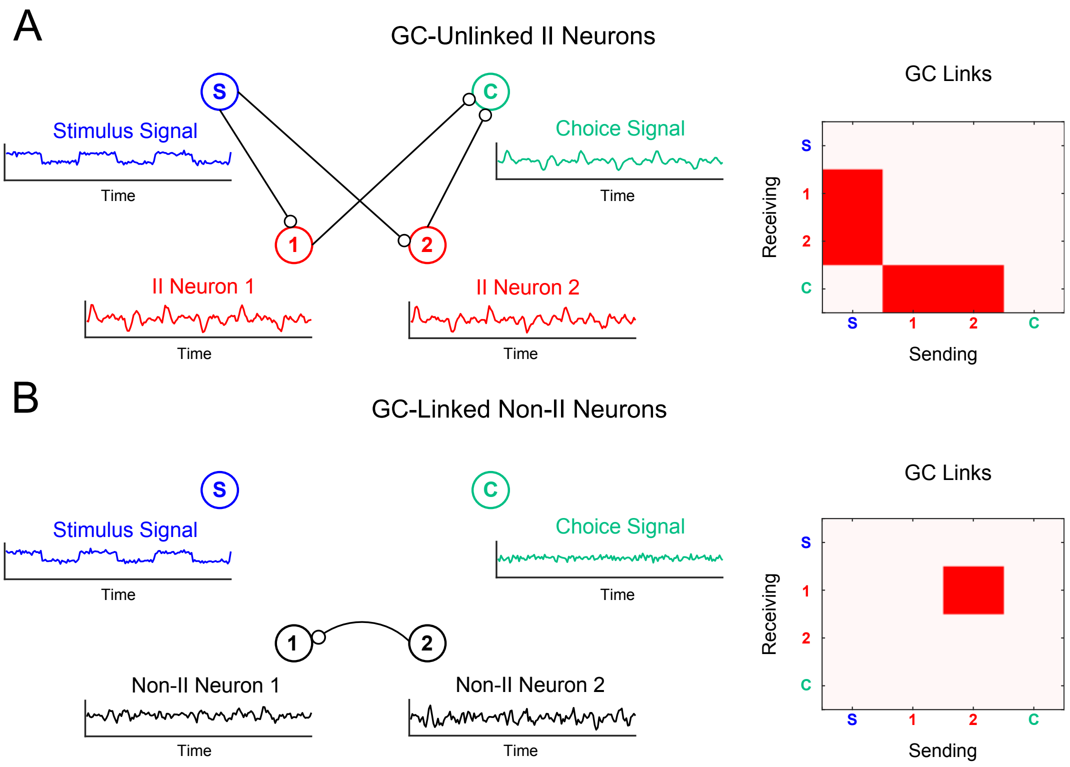
Nikolas A. Francis[1,‡], Shoutik Mukherjee[3,‡], Loren Koçillari[4,5,‡], Stefano Panzeri[4,5,*], Behtash Babadi[3,*], Patrick O. Kanold[1,2,*,†]

Supplemental Figures S1-S6
Supplemental Tables S1-S6

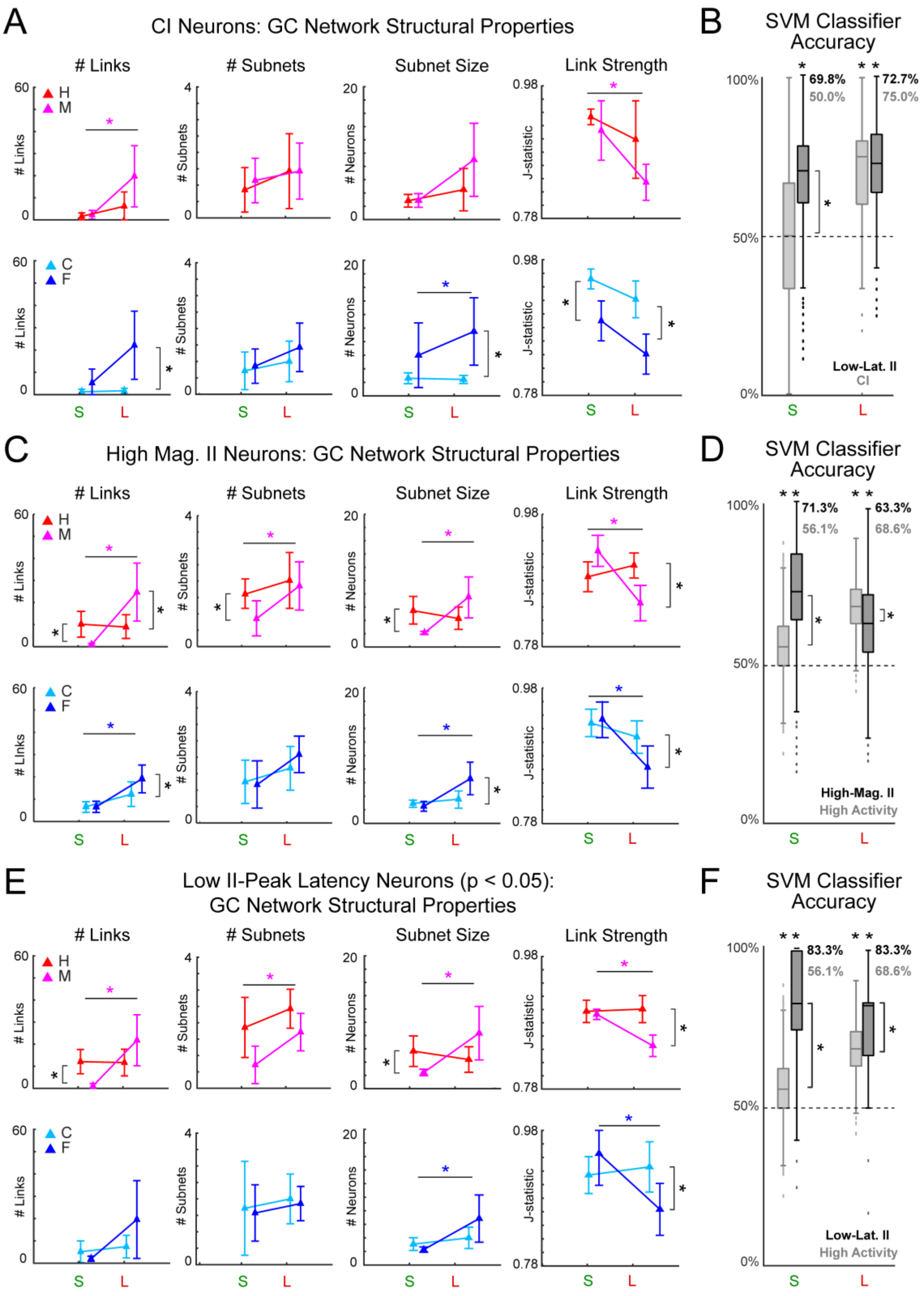
SUPPLEMENTAL FIGURES



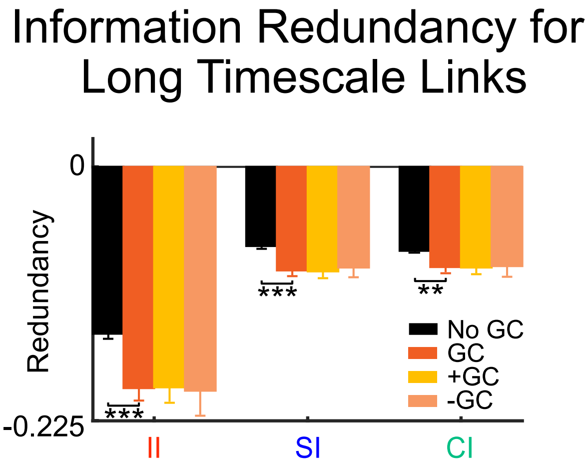
Supplemental Figure S1. Time-course of *SI*, *CI*, and *II* averaged over neurons that carried stimulus information (*SI*) only in **A**-**B** and choice information (*CI*) only in **C**-**D**. As in Figure 3, we quantified the *SI*, *CI*, and *II* in six separate stages of the behavioral task, which account for the peri-stimulus (0-1.5 s) and the post-stimulus intervals (1.5-3 s) shown by the shaded regions. Error bars show one standard error of the mean (SEM; N=#neurons with SI, CI peaks within the stage).



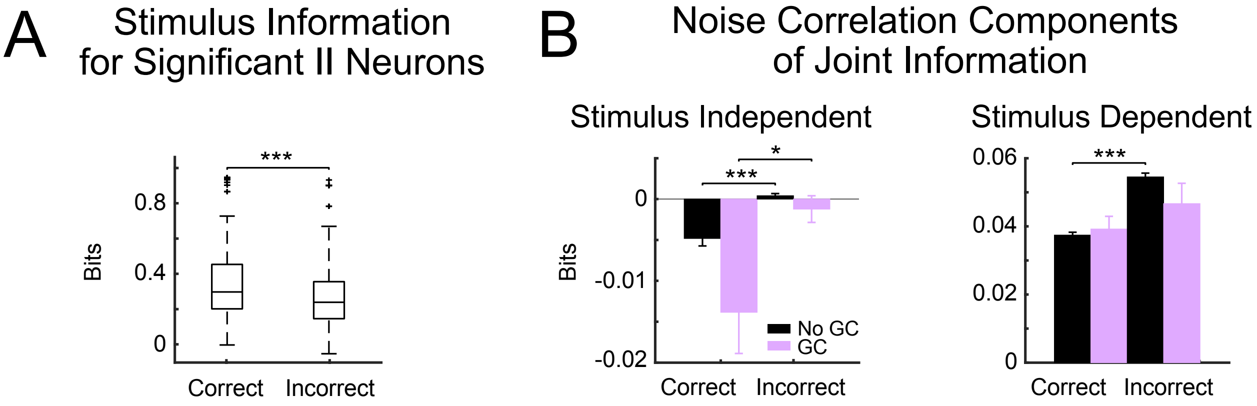
Supplemental Figure S2. Two neurons having *II* is neither necessary nor sufficient for them to be GC-linked. Related to Figure 4. **A.** The first case shows two *II* neurons, i.e. two neurons that are modulated by stimulus signal and both modulate a choice signal. However, they do not modulate each other's activity and hence there is no GC link. **B.** The second case shows two neurons in which the activity of neuron 1 is modulated by the activity of neuron 2, i.e. there is a GC link from neuron 2 to neuron 1. However, neither neuron is modulated by the stimulus signal nor do they modulate the choice signal and so the two neurons do not have *II*.



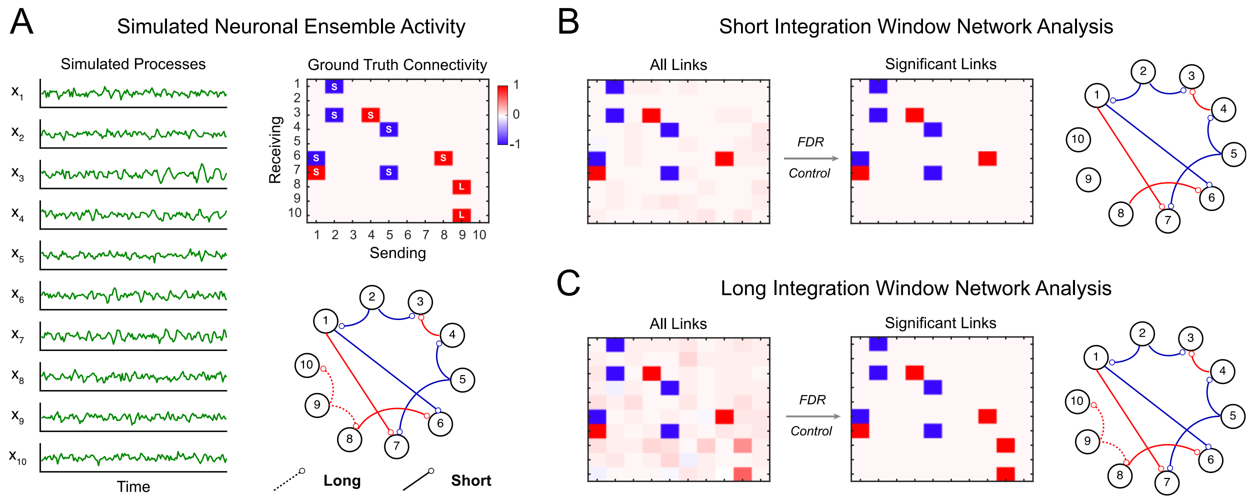
Supplemental Figure S3. Network structure of neurons with *CI*, but not *II*, neurons with greatest peak *II* magnitudes, and neurons with lowest *II*-peak latency but with more strict selection threshold. Related to Figure 4. **A.** Functional networks of short (S)- and long (L)- timescale interactions amongst neurons with high *CI*, but not *II*, were estimated using GC analysis for each behavioral choice. 20 exclusively *CI* neurons with above-threshold peak *CI* and sub-threshold peak *II* that had the shortest *CI* peak latency were selected for GC network analysis in each session with at least 20 *CI* neurons. Number of links, number of subnetworks, size of subnetworks, and statistical strength of links are represented as means ± 2 SEM. Asterisks indicate statistically significant differences based on Wilcoxon’s signed rank test ($p<0.05$). See also Supp. Table S3. **B.** Network statistics were used to train an SVM to classify into correct or incorrect decisions. Across timescale and selection of neurons—except *CI* S-timescales—decisions were predicted significantly better than chance ($p<0.001$). S-timescale network structure of *CI* neurons was decoded at chance-level accuracy, less than of low-latency *II* neurons ($p<0.001$), shown in Fig. 4. L-timescale network structure had higher decoding accuracy, and *CI* neuronal networks were decoded with similar accuracy ($p=0.708$). A two-sample t-test ($p<0.05$) was used to compare to neurons with low-latency *II*, and a one-sample t-test ($p<0.05$) to compare performance with chance decoding accuracy. **C.** Network statistics of greatest peak *II* magnitude neurons are shown in the same format as in panel A. Network statistics differed by timescale and behavioral choice similarly to network statistics of low *II*-peak latency neurons (Figure 4C). See also Supplemental Table S4. **D.** Network statistics were used to train an SVM to classify into correct or incorrect decisions. Across timescale and selection of neurons, decisions were predicted significantly better than chance ($p<0.001$). S-timescale network structure of high-magnitude *II* neurons was better decoded than of highly responsive neurons ($p<0.001$). L-timescale network structures had high decoding accuracy, but highly responsive neuronal networks were better decoded ($p<0.001$). **E.** Network statistics of low *II*-peak latency neurons chosen with *II* threshold of $p<0.05$ (see methods) are shown in the same format as in panel A. Network statistics differed by timescale and behavioral choice similarly to network statistics of low *II*-peak latency neurons (Figure 4C). Network structure of neurons with lowest *II*-peak latency is robust to selection threshold. See also Supplemental Table S5. **F.** Network statistics were used to train an SVM to classify into correct or incorrect decisions. Across timescale and selection of neurons, decisions were predicted significantly better than chance ($p<0.001$). S-timescale network structure of low *II*-peak latency neurons was better decoded than of highly responsive neurons ($p<0.001$). L-timescale network structures had high decoding accuracy, and low *II*-peak latency neuronal networks were still better decoded ($p<0.001$). Thus, network structure of low *II*-peak latency neurons is robust to choice of *II* threshold. Importantly, decoding accuracies are higher (83.3%) with the lower threshold than our main result in Fig. 4C, but distributions are more skewed. Asterisks indicate statistically significant differences. Wilcoxon's signed rank test ($p<0.05$) was used in panels **A**, **C**, and **E**, a two-sample t-test ($p<0.05$) to compare to neurons with high activity, and a one-sample t-test ($p<0.05$) to compare performance with chance decoding accuracy in panels **B**, **D,** and **F**.



Supplemental Figure S4. The normalized time-lagged redundancy index showed that GC-linked pairs of neurons in L-timescale networks shared more redundant information (*II, SI, CI*) than GC-unlinked pairs of neurons (see Figure 6 in the main text for a comparison to S-timescale networks). Related to Figure 6. We report no difference in redundancy index between groups of positive and negative GC-linked neurons. Statistical comparisons between groups are made with a two-sample t-test (*p<0.05; **p<0.01; ***p<0.001).



Supplemental Figure S5. Stimulus information of significant *II* neurons and contribution of noise correlation to joint information of neural pairs during correct and incorrect behavior. Related to Figure 6. **A.** Stimulus information was computed separately for correct and incorrect trials for significant *II* neurons at the time peak of each neuron's information. Stimulus information was higher in correct trials. The upper and lower edges of the boxes show the 75th and 25th percentiles, respectively, and the horizontal line marks the sample median. **B.** Using the information breakdown approach of (Pola et al., 2003), we broke down the contribution of noise correlations to the joint information carried by neuron pairs into the finer sub-components of stimulus-independent (left panel) and stimulus-dependent (right panel) noise correlations, separately for GC-linked and GC-unlinked pairs. Stimulus-dependent information decreased in incorrect trials for both groups of cells, while stimulus-dependent information increased in incorrect trials for GC-unlinked pairs. Bar plots show mean ± SEM. Results are reported in bits (as opposed to Fig. 6 of the main text where we plotted normalized redundancy/synergy values). To avoid systematic error (bias) in the estimation of information due to the different number of correct and error trials, we equalized the number of correct and incorrect trials by randomly subsampling the correct trials. Asterisks indicate statistically significant differences computed with a Wilcoxon's signed rank test (*p<0.05, ***p<0.001). The synergy between neurons during incorrect trials shown in Fig. 6D, despite having positive signal and noise correlations, is due to the major decrease of signal correlations, through the deterioration of the stimulus information (panel **A**), and the decrease of noise correlation strengths. Panel **B** suggests that the decrease of signal and noise correlation strengths greatly diminished the information-limiting effects of stimulus-independent correlations and left only the information enhancing effect of stimulus-dependent correlations, which led to synergy.



Supplemental Figure S6. Simulated example for assessing the use of the proposed parametrization of the integration window lengths in Granger causality (GC) network inference. Related to STAR Methods. **A.** Simulated responses of 10 neurons, shown averaged over 10 trials of 150 time samples in the left panel, were generated based on an underlying network of long (L) and short (S) timescale interactions (right subpanels). **B.** GC analysis using the short integration window identifies true S-timescale interactions, while expectedly discarding the L-timescale influences. False discovery rate (FDR) control prunes weak spurious interactions and retains significant links. **C.** Employing the L integration window for GC analysis captures both S and L influences, and after FDR control, the true functional connectivity is inferred correctly.

SUPPLEMENTAL TABLES

	# Links			# Subnets			Subnet Size			Link Strength		
	S	L	p-value	S	L	p-value	S	L	p-value	S	L	p-value
H	8.25 ± 1.61	8.33 ± 1.49	0.97	1.58 ± 0.19	2.08 ± 0.31	0.19	5.11 ± 0.75	4.16 ± 0.61	0.333	0.91 ± 0.009	0.89 ± 0.006	0.111
M	2.08 ± 1.05	21.42 ± 6.10	**0.009**	0.83 ± 0.24	1.50 ± 0.26	0.074	3.30 ± 0.99	7.94 ± 1.91	**0.041**	0.91 ± 0.012	0.85 ± 0.014	**0.014**
p-value	**0.005**	0.059		**0.024**	0.167		0.161	0.073		0.975	**0.02**	
C	2.58 ± 0.91	4.58 ± 2.32	0.189	1.00 ± 0.28	1.92 ± 0.34	**0.047**	3.08 ± 0.42	3.00 ± 0.43	0.59	0.95 ± 0.011	0.93 ± 0.015	0.369
F	2.33 ± 0.73	20.5 ± 3.90	**<0.001**	1.42 ± 0.31	1.50 ± 0.19	0.824	2.59 ± 0.32	10.0 ± 1.72	**<0.001**	0.91 ± 0.012	0.85 ± 0.016	**0.013**
p-value	0.832	**0.002**		0.33	0.298		0.357	**<0.001**		**0.049**	**0.002**	

Supplemental Table S1. Statistical comparisons of GC network structure across short (S) and long (L) timescales, and behavioral choice categories — hit (H), miss (M), correct-rejection (C), and false-alarm (F) – using Wilcoxon’s signed rank test (p<0.05). See also Fig. 4C.

	# Links			# Subnets			Subnet Size			Link Strength		
	S	L	p-value	S	L	p-value	S	L	p-value	S	L	p-value
H	8.41 ± 1.605	13.91±2.885	0.102	1.63 ± 0.178	2.53 ± 0.229	**0.003**	4.64 ± 0.556	4.48 ± 0.486	0.836	0.91 ± 0.009	0.88 ± 0.008	**0.023**
M	1.88 ± 0.453	29.91±4.544	**<0.001**	1.13 ± 0.233	1.47 ± 0.168	0.235	2.58 ± 0.171	10.11±1.174	**<0.001**	0.92 ± 0.007	0.84 ± 0.009	**<0.001**
p-value	**<0.001**	**0.004**		0.093	**<0.001**		**<0.001**	**<0.001**		0.457	**<0.001**	
C	6.72 ± 1.112	12.31±2.817	0.072	2.06 ± 0.206	2.34 ± 0.236	0.372	3.67 ± 0.390	4.48 ± 0.467	0.183	0.92 ± 0.007	0.90 ± 0.008	**0.013**
F	6.63 ± 1.261	19.25±3.057	**<0.001**	1.88 ± 0.245	1.59 ± 0.148	0.33	3.90 ± 0.454	8.23 ± 0.922	**<0.001**	0.90 ± 0.009	0.86 ± 0.009	**0.002**
p-value	0.956	0.1		0.56	**0.01**		0.697	**<0.001**		0.051	**0.001**	

Supplemental Table S2. Highly responsive neurons: statistical comparisons of GC network structure across short (S) and long (L) timescales, and behavioral choice categories — hit (H), miss (M), correct-rejection (C), and false-alarm (F) – using Wilcoxon's signed rank test ($p<0.05$). See also Figure 4.

	# Links			# Subnets			Subnet Size			Link Strength		
	S	L	p-value	S	L	p-value	S	L	p-value	S	L	p-value
H	1.57 ± 0.783	6.14 ± 3.233	0.213	0.86 ± 0.340	1.43 ± 0.572	0.411	2.83 ± 0.477	4.50 ± 1.594	0.339	0.93 ± 0.006	0.90 ± 0.029	0.385
M	2.29 ± 0.993	19.71±6.925	**0.046**	1.14 ± 0.340	1.43 ± 0.429	0.612	2.88 ± 0.516	9.0 ± 2.749	0.054	0.91 ± 0.022	0.83 ± 0.014	**0.04**
p-value	0.583	0.116		0.563	1		0.954	0.178		0.492	0.142	
C	1.29 ± 0.566	1.71 ± 0.566	0.602	0.71 ± 0.286	1.0 ± 0.309	0.51	2.60 ± 0.400	2.43 ± 0.297	0.74	0.95 ± 0.007	0.92 ± 0.014	0.157
F	5.29 ± 3.08	22.14±7.640	0.075	0.86 ± 0.261	1.43 ± 0.369	0.233	6.0 ± 2.381	9.50 ± 2.487	**0.033**	0.89 ± 0.015	0.84 ± 0.015	0.069
p-value	0.246	**0.037**		0.718	0.391		0.215	**0.019**		**0.021**	**0.006**	

Supplemental Table S3. Neurons exclusively with *CI*: statistical comparisons of GC network structure across short (S) and long (L) timescales, and behavioral choice categories — hit (H), miss (M), correct-rejection (C), and false-alarm (F) – using Wilcoxon's signed rank test ($p<0.05$). See also Figure 4 and Supplemental Figure S3A.

	# Links			# Subnets			Subnet Size			Link Strength		
	S	L	p-value	S	L	p-value	S	L	p-value	S	L	p-value
H	10.17±2.915	8.92 ± 2.704	0.756	2 ± 0.229	2 ± 0.427	0.401	5.42 ± 1.050	4.25 ± 0.813	0.383	0.89 ± 0.011	0.90 ± 0.001	0.282
M	0.92 ± 0.286	24.67 ± 6.59	**0.004**	0.83 ± 0.271	1.83 ± 0.366	**0.04**	2.1 ± 0.10	7.46 ± 1.036	**0.002**	0.93 ± 0.012	0.85 ± 0.013	**0.002**
p-value	**0.009**	**0.043**		**0.046**	0.77		**0.005**	0.074		0.065	**0.002**	
C	3.17 ± 1.10	6.17 ± 1.714	0.157	1.25 ± 0.329	1.67 ± 0.334	0.383	2.93 ± 0.284	3.55 ± 0.651	0.393	0.93 ± 0.011	0.91 ± 0.012	0.27
F	1.83 ± 0.695	16.17±4.114	**0.005**	1.17 ± 0.366	2 ± 0.288	0.062	2.5 ± 0.360	6.56 ± 1.221	**0.003**	0.93 ± 0.014	0.86 ± 0.015	**0.009**
p-value	0.319	**0.041**		0.867	0.354		0.353	**0.036**		0.798	**0.035**	

Supplemental Table S4. Neurons with greatest *II*-peak magnitudes: statistical comparisons of GC network structure across short (S) and long (L) timescales, and behavioral choice categories — hit (H), miss (M), correct-rejection (C), and false-alarm (F) – using Wilcoxon’s signed rank test ($p<0.05$). See also Figure 4 and Supplemental Figure S3C.

	# Links			# Subnets			Subnet Size			Link Strength		
	S	L	p-value	S	L	p-value	S	L	p-value	S	L	p-value
H	12.0 ± 2.76	11.6 ± 3.02	0.918	1.86 ± 0.46	2.43 ± 0.30	0.32	5.62 ± 1.15	4.35 ± 0.96	0.408	0.90 ± 0.008	0.90 ± 0.010	0.801
M	1.00 ± 0.53	21.7 ± 5.78	**0.012**	0.71 ± 0.29	1.71 ± 0.29	**0.029**	2.40 ± 0.24	8.33 ± 2.02	**0.014**	0.89 ± 0.004	0.85 ± 0.008	**<0.001**
p-value	**0.007**	0.154		0.061	0.11		**0.017**	0.094		0.659	**0.001**	
C	5.14 ± 2.41	7.43 ± 2.53	0.526	1.71 ± 0.71	2.00 ± 0.38	0.732	3.08 ± 0.47	4.00 ± 0.78	0.327	0.91 ± 0.14	0.93 ± 0.019	0.654
F	2.00 ± 0.58	19.6 ± 8.72	0.091	1.57 ± 0.43	1.86 ± 0.26	0.582	2.27 ± 0.20	6.85 ± 1.74	**0.022**	0.95 ± 0.023	0.86 ± 0.019	**0.019**
p-value	0.247	0.223		0.867	0.762		0.131	0.154		0.311	**0.037**	

Supplemental Table S5. Neurons with low *II*-peak latency (threshold p<0.05): statistical comparisons of GC network structure across short (S) and long (L) timescales, and behavioral choice categories — hit (H), miss (M), correct-rejection (C), and false-alarm (F) – using Wilcoxon's signed rank test (p<0.05). See also Figure 4 and Supplemental Figure S3E.

	# Links		# Subnets		Subnet Size		Link Strength	
	S	L	S	L	S	L	S	L
Low-Latency II Networks								
H vs. C	**0.007**	0.06	0.098	0.72	**0.025**	0.128	**0.0143**	**0.017**
M vs. F	0.85	0.901	0.155	1	0.508	0.429	0.803	0.769
Highly Responsive Networks								
H vs. C	0.391	0.694	0.113	0.57	0.157	0.998	0.277	0.229
M vs. F	**0.001**	0.055	**0.03**	0.579	**0.008**	0.213	0.125	0.091
CI Networks								
H vs. C	0.773	0.223	0.753	0.525	0.717	0.231	0.196	0.582
M vs. F	0.384	0.818	0.518	1	0.251	0.894	0.488	0.797
High Magnitude II Networks								
H vs. C	**0.041**	0.401	0.415	0.545	**0.033**	0.505	**0.033**	0.868
M vs. F	0.242	0.288	0.472	0.597	0.301	0.651	0.771	0.464

Supplemental Table S6. Supplemental statistical comparisons of correct/incorrect decision GC network structure by stimulus — i.e., hit (H) vs. correct-rejection (C), and miss (M) vs. false-alarm (F) – for short (S) and long (L) timescales using Wilcoxon’s signed rank test ($p<0.05$). Comparisons performed for networks of low *II*-peak latency neurons (see also Figure 4 and Table S1); highly responsive neurons (see also Supp. Table S2); *CI* neurons (see also Supp. Fig. S3A-B and Table S3); and high-magnitude *II* neurons (see also Supp. Fig. S3C-D and Supp. Table S4).