# Audiovisual task switching rapidly modulates sound encoding in mouse auditory cortex

**Author list:**
Ryan J. Morrill[1,2,3]; James Bigelow[1,3]; Jefferson DeKloe[1,3]; Andrea R. Hasenstaub[1,2,3,*]

[1] Coleman Memorial Laboratory
[2] Neuroscience Graduate Program
[3] Department of Otolaryngology–Head and Neck Surgery, University of California, San Francisco, 94143
* Corresponding author

Nr. figs: 8
Nr. supplemental figs. / tables: 4 / 11

**Keywords:**
attention; crossmodal; multisensory; behavior; cortical layers; extracellular physiology

**Acknowledgements:**
This work was supported by National Institutes of Health grants R01DC014101 and NS116598 to A.R.H., the National Science Foundation GRFP to R.J.M., the Klingenstein Foundation to A.R.H., Hearing Research Inc. to A.R.H., and the Coleman Memorial Fund to A.R.H.

# Abstract

In everyday behavior, sensory systems are in constant competition for attentional resources, but the cellular and circuit-level mechanisms of modality-selective attention remain largely uninvestigated. We conducted translaminar recordings in mouse auditory cortex (AC) during an audiovisual (AV) attention shifting task. Attending to sound elements in an AV stream reduced both pre-stimulus and stimulus-evoked spiking activity, primarily in deep layer neurons and neurons without spectrotemporal tuning. Despite reduced spiking, stimulus decoder accuracy was preserved, suggesting improved sound encoding efficiency. Similarly, task-irrelevant mapping stimuli during intertrial intervals evoked fewer spikes without impairing stimulus encoding, indicating that attentional modulation generalized beyond training stimuli. Importantly, spiking reductions predicted trial-to-trial behavioral accuracy during auditory attention, but not visual attention. Together, these findings suggest auditory attention facilitates sound discrimination by filtering sound-irrelevant background activity in AC, and that the deepest cortical layers serve as a hub for integrating extramodal contextual information.


# Introduction

Information from one or another sensory pathway may become differentially relevant due to environmental changes. The brain must therefore continuously assign limited attentional resources to processing simultaneous information streams from each sensory modality. For example, hearing a siren while listening to music in the car might prompt an attention shift away from the auditory stream, toward a visual search for emergency vehicles. On the other hand, a similar shift away from the music is unlikely while listening at home. In these cases, contextual cues support allocating attention to either the auditory domain or the visual domain, and the perceptual experience of the music is qualitatively different. How might sensory cortex differentially encode stimuli from an attended versus filtered modality?

Attentional selection operates cooperatively at many levels of sensory processing. Most effort has been devoted to understanding the neural mechanisms of feature-selective attention within a single modality (Desimone and Duncan, 1995; Fritz et al., 2007). A major focus of this work has been characterizing transformations of stimulus representations in sensory cortical areas, due to their pivotal position between ascending sensory pathways and behavioral networks implementing top-down control (Lamme et al., 1998; Sutter and Shamma, 2011). These studies, largely from the visual domain, have shown that attention to a stimulus feature or space will often increase stimulus-evoked spiking responses and reduce thresholds for eliciting

a response; likewise, responses to unattended stimuli are often decreased (Reynolds and Chelazzi, 2004). On the other hand, fewer studies have examined how modality-selective attention affects encoding in sensory cortex. This mode of attention highlights behaviorally-relevant sensory streams while filtering less relevant ones. Human fMRI studies have reported differential activation patterns in auditory and visual cortex (AC, VC) reflecting the attended modality (Johnson and Zatorre, 2005; Petkov et al., 2004; Shomstein and Yantis, 2004; Woodruff et al., 1996). Extending these findings, studies in primate AC and VC have reported entrainment local field potential (LFP) oscillations by modality-selective attention, which serves to modulate excitability and sharpen feature tuning within sensory cortex corresponding to the attended modality (Hocherman et al., 1976; Lakatos et al., 2009, 2008; O'Connell et al., 2014). Several findings suggest these influences may differ among cortical layers and between inhibitory and excitatory neurons (Lakatos et al., 2016; O'Connell et al., 2014).

Nevertheless, much remains unknown about the influence of modality-specific attention on stimulus encoding in sensory cortex. Importantly, potential interplay between ongoing activity and evoked responses, as well as their consequences for information and encoding efficiency, has not been examined. How these influences may be differentially expressed in cell subpopulations defined by cortical depth or inhibitory/excitatory cell type similarly remains unknown. Finally, the degree to which influences of modality-specific attention may generalize beyond training stimuli remains unknown.

In the present study, we addressed these open questions by examining single neuron activity and sensory responses in mouse AC during an audiovisual attention shifting task. AC integrates ascending auditory information with diverse input from frontal, cingulate, striatal, and non-auditory sensory areas to rapidly alter sensory processing in response to changing behavioral demands (Budinger et al., 2008; Budinger and Scheich, 2009; Park et al., 2015; Rodgers and DeWeese, 2014; Winkowski et al., 2013). To isolate the influence of modality-selective attentional modulation, we compared responses to identical compound auditory-visual stimuli within task, under different cued contexts requiring attention to the auditory or visual elements, thus holding constant other task-related variables such as arousal, attention, reward expectation, and motor activity (Saderi et al., 2021). Because spike rate and information changes are dissociable (Bigelow et al., 2019; Phillips and Hasenstaub, 2016), we quantified both evoked spike rates and the mutual information between responses and stimuli. We also examined the generality of modality-specific attention by examining responses to task-irrelevant sounds presented between trials. Finally, we used translaminar probes and spike waveform

morphology classification to capture possible attention-related differences in neurons among cortical layers and between putative inhibitory and excitatory cell classes.

# Results

***Audiovisual rule-switching in mice.*** We trained mice to perform an AV rule-switching task, in which they made decisions using auditory stimuli while ignoring simultaneously presented visual stimuli or vice versa. Trial presentation was self-paced in a virtual foraging environment wherein a visual track was advanced by forward locomotion on a spherical treadmill (**Figure 1.A**). A task-irrelevant random double sweep (RDS) sound was presented during intertrial intervals (ITIs) for mapping auditory receptive fields in each attentional state (**Figure 1.B**). Decision stimuli were presented after variable track length, comprising 1-s auditory tone clouds (TCs; centered at 8 or 17 kHz) and/or visual drifting gratings (horizontal or vertical orientation; **Figure 1.C**). One of the decision stimuli for each modality was a rewarded target ($A_R$, $V_R$) and the other an unrewarded distractor ($A_U$, $V_U$). Lick responses following targets (hits) and distractors (false alarms; FAs) produced water rewards and dark timeouts, respectively. Withholding licks for targets (misses) or distractors (correct rejects; CRs) advanced the next trial. Each session began with a block of unimodal decision stimuli, which cued the attended modality of a subsequent AV block (**Figure 1.D**). A second unimodal block from the other modality was then presented, cueing the rule for a final AV block. Decision stimuli had identical physical properties but different behavioral significance between rules (e.g., licks following $A_RV_U$ were rewarded in A-rule but punished in V-rule). Targets and distractor stimuli remained constant throughout training for each mouse and were approximately counterbalanced across animals. Block sequences (A-rule then V-rule, or vice versa) were also counterbalanced across sessions (**Figure 1.D.c**).

We used two approaches to ensure that animals were engaged during both task rules. First, we restricted analysis to sessions in which discrimination was well above chance (*d'* >1.5) for both rules, and for which FA rates were below 0.5 for the stimuli with reward valences that conflicted across rules ($A_UV_R$ in the A-rule, $A_RV_U$ in the V-rule; **Figure 1.F**). Second, for a subset of sessions (n = 14 sessions, 5 mice) we measured pupil size, a well-established correlate of arousal and behavioral performance (Bradley et al., 2008; McGinley et al., 2015; Reimer et al., 2014). We used a computer vision algorithm to automate measurement of pupil size (pupil diameter / eye diameter) for each frame acquired by a CCD video camera (**Figure 2.A.a**). To isolate pupil fluctuations reflecting general arousal, pupil size was measured during an inter-trial

interval (ITI) window designed to avoid pupil responses to decision stimulus onset, dark timeouts, and decreased locomotion events following reward administration (**Figure 2.A.b-c, Figure 2.B**). Previous studies have reported that pupil size increases with task difficulty and engagement in humans, non-human primates and rodents (Hess and Polt, 1964; Kawaguchi et al., 2018; Schriver et al., 2018). We reasoned that comparison of pupil size across the rules would allow us to establish whether task demands differed between the rules. No difference in pupil size was observed between rules during bimodal blocks (**Figure 2.C**; A-rule bimodal: 0.29 ±0.05 norm. diameter ± SD, V-rule bimodal: 0.30 ± 0.05; Z = -1.0, p = 0.30, paired Wilcoxon signed rank [WSR], Benjamini-Hochberg false discovery rate [FDR]-adjusted *p*-values). As expected, pupil diameters were significantly larger during the bimodal portion of the task, when visual stimuli were present and the task had increased in difficulty, compared to the auditory-only unimodal portion of the task (**Figure 2 - figure supplement 1;** A-rule unimodal: 0.28 ±0.04, A-rule bimodal: 0.29 ±0.05; Z = -2.6, p = 0.009). A trend toward smaller pupil size in the unimodal visual rule compared to bimodal rule was also noted, but the difference did not reach significance after multiple comparisons correction (V-rule unimodal: 0.29 ±0.04, V-rule bimodal: 0.30 ±0.05; Z = -2.0, p = 0.062). Because pupil size also closely tracks locomotion (**Figure 2.A.b-c**; McGinley et al., 2015), we examined locomotion speed during the same ITI window (**Figure 2.D**). Differences in locomotion speed were also not observed between rules (all p ≥ 0.623; all |z| ≤ 1.29, paired WSR). Arousal and motor activity were thus comparable between rules, suggesting that differences in neuronal activity may be attributable to modality-selective attention.

***Single unit recording in AC.*** After mice learned the AV rule-switching task, a craniotomy was made over right AC, to allow for acute recordings during behavior using multichannel probes spanning the full cortical depth (**Figure 3.A**). In total, we recorded AC activity in 10 mice during 23 behavioral sessions meeting inclusion criteria. The putative cortical depth of each sorted single unit (SU) was assigned by calculating the fractional position of the channel with the largest waveform amplitude within the span of channels in AC as estimated from spontaneous and tone-evoked recordings following the task (**Figure 3.B**). A separate set of experiments using probe track visualization with Di-I provided support for this depth estimation technique (**Figure 3.C**; (DiCarlo et al., 1996; Morrill and Hasenstaub, 2018). We then divided the fractional depth values into superficial, middle, and deep groups, approximating the supragranular, granular, and infragranular laminae. We further divided SUs into narrow-spiking (NS, putative inhibitory; n = 130, 18%) and broad-spiking (BS, predominantly excitatory; n = 612, 82%)

populations based on trough-peak time (**Figure 3.D**; Bigelow et al., 2019; Cardin et al., 2007; Nandy et al., 2017; Phillips et al., 2017).

***Modality-selective attention modulates stimulus-evoked firing rates.*** Our analysis of SU activity during AV rule-switching focused on comparisons of responses to bimodal decision stimuli between rules, which reflected physically identical stimuli and comparable arousal and locomotion levels. Responses evoked by decision stimuli and modulatory effects of task rule were diverse (**Figure 4.A**). To capture a predominantly sensory-driven component of the response, we measured mean firing rates (FRs) during the first 300 ms post-stimulus onset (**Figure 4.B**), which preceded most lick responses (lick latency median: 611 ms; 5th-95th percentiles: 289-1078 ms; 5.4% of licks <300 ms, n = 2,852 total lick trials across dataset). Trials with licks earlier than 300 ms were excluded from analysis. We first compared A-rule and V-rule responses to the tone cloud rewarded in the A-rule ($A_R$*: $A_RV_R$ and $A_RV_U$ responses combined). Averaging across units, responses in the deep layers were suppressed in the A-rule relative to the V-rule for both NS and BS units (**Figure 4.C** Deep. BS: p = 2.8e-4, Z = 4.1, median fold change [FC; A-rule/V-rule]: 0.89, n = 333 SUs; Deep NS: p = 0.011, Z = 2.9, med. FC: 0.87, n = 66; paired WSR, FDR-adjusted *p*-values; see **Figure 4 - source data 1.A** for full stats). No significant group-level change was found in middle or superficial units. Consistent with group level trends, individual units with significant FR decreases in the A-rule ($p < 0.01$, unpaired t-test) substantially outnumbered units with significant FR increases for all unit populations other than superficial and mid-depth BS units (**Figure 4.C**, right).

A similar pattern of attention-related modulation was observed for unrewarded stimuli in the A-rule ($A_U$*: $A_UV_R$ and $A_UV_U$ responses combined). At the group level, superficial and middle unit responses did not significantly differ between conditions, whereas deep BS units were significantly suppressed in the A-rule (**Figure 4.D**; Deep BS: p = 2.0e-06, Z = 5.11, med. FC: 0.81, n = 321; paired WSR, FDR-adjusted *p*-values; see **Figure 4 - source data 1.B** for full stats). Relative fractions of units with significantly modulated FRs to $A_U$* stimuli were similar to those described above for $A_R$* stimuli (**Figure 4.D**, right), with the exception of the superficial group, in which slightly more units were significantly increased. We further found that most units showed the same direction of modulation for $A_R$* and $A_U$* stimuli (**Figure 4.E**), with similar modulation sign observed for 78% of BS units (50% suppressed for both $A_R$* and $A_U$*, 28% enhanced for both) and 99% of NS units (68% suppressed for both, 31% enhanced for both). These findings suggest that modality-selective attention similarly influences FRs evoked by task-relevant target and distractor sounds with different acoustic properties and learned behavioral values.

To determine whether these attentional influences might generalize to task-irrelevant sounds, we examined responses to RDS sounds presented during the ITI. Using the same analysis window (300 ms post-stimulus onset, **Figure 4.F**), we found that attention-related modulation of FR responses evoked by task-irrelevant sounds was highly similar to that observed for both types of decision stimuli: middle- and deep-layer BS and NS populations exhibited group-level FR suppression during the A-rule (**Figure 4.G**), whereas superficial layer units were not significantly modulated (Mid. BS: p = 6.0e-3, Z = 3.1, med. FC: 0.85, n = 112; Mid. NS: p = 0.014, Z = 2.7, med. FC: 0.65, n = 28; Deep BS: p = 1.2e-6, Z = 5.2, med. FC: 0.84, n = 309; Deep NS: p =0.021, Z = 2.5, med. FC: 0.80, n = 64; paired WSR; see **Figure 4 - source data 2** for full stats). Significantly modulated unit counts were again highly biased towards suppression in the A-rule, with pronounced differences in the middle and deep unit groups (**Figure 4.G**, right). Together, these results show that auditory-selective attention tends to reduce FR responses to sounds, regardless of their behavioral relevance, valence, or spectral content, and that these influences are strongest for deep layer units.

***Modality-selective attention also modulates pre-stimulus firing rates.*** Previous studies have found that modulation of ongoing activity in sensory cortex can influence subsequent sensory-evoked responses (Arieli et al., 1996; Haider and McCormick, 2009). Thus, the response suppression during auditory attention reported above may either reflect specific decreases in stimulus responsivity or general decreases in ongoing activity. To address these possibilities, we quantified FRs in a pre-stimulus window spanning 300 ms prior to decision stimulus onset in which no sounds were presented (**Figure 5.A.a**). Although this window may include anticipatory modulation of activity (Cox et al., 2019; Egner et al., 2010; Samuelsen et al., 2012), it nevertheless provides a measure of baseline activity for comparison with evoked responses. We observed significant group-level decreases in pre-stimulus FRs during the A-rule for units in the middle NS and deep BS groups, but no modulation of superficial units (**Figure 5.B**; Mid. NS: p = 0.039, Z = 2.48, med. FC: 0.71, n = 28; Deep BS: p = 4.2e-05, Z = 4.49, med. FC: 0.87, n = 336; paired WSR, FDR-adjusted *p*-values; see **Figure 5 - source data 1** for full stats). To test whether the reduction in pre-stimulus FR was sufficient to account for stimulus-evoked changes reported above, we recalculated FRs evoked by decision stimuli as fold change from pre-stimulus FRs (**Figure 5.A.b**). Following this adjustment and after false discovery rate correction, the middle- and deep-layer unit population responses no longer differed between rules (**Figure 5.C; Figure 5 - source data 2**). Together, these results suggest that group level decreases in evoked FRs during A-rule are largely due to generalized suppression of ongoing AC activity.

***Attention-related suppression is driven by units without STRF tuning.***

We next sought to determine whether attention-related changes in stimulus response were related to the tuning preferences of units, a phenomenon termed “feature attention” previously observed in both monkey visual cortex (Maunsell and Treue, 2006; Treue and Martínez Trujillo, 1999) and AC (Da Costa et al., 2013). The RDS mapping stimulus, which we have previously used to efficiently identify auditory response properties and audiovisual interactions (Bigelow et al., 2022), was used to generate spectrotemporal receptive fields (STRFs) through reverse correlation (**Figure 6.A;** Aertsen and Johannesma, 1981; de Boer, 1968; Gourévitch et al., 2015). Tuning for each STRF was measured through a trial-to-trial reliability metric, which we used to divide units into those with activity changes that were reliably evoked by defined spectral or temporal features (tuned STRFs, n = 172; **Figure 6.C.a**) and those without feature-evoked changes (untuned, n = 409; **Figure 6.C.b**). Spiking activity levels were higher in tuned units compared to untuned (**Figure 6.D.a**). To control for possible activity level-dependent effects, we compared our population of tuned units to a randomly selected subset of untuned units which was matched for both sample size and FR to the tuned population (**Figure 6.D.b**). We then compared attentional modulation of stimulus responses between the tuned and untuned groups. Responses to the rewarded TC ($A_R$*), unrewarded TC ($A_U$*) and the RDS mapping stimuli were significantly modulated by task rule in the untuned group, but not the tuned group (tuned: all $p \geq 0.18$, all $|Z| \leq 1.72$; untuned: all $p \leq 0.023$, all $|Z| \geq 2.27$; one-way WSR vs. modulation of 1 [equal across rules], FDR-adjusted *p*-values; **Figure 6 - source data 1.C,D**) Nevertheless, comparisons across these tuned and untuned groups showed that the distributions did not significantly differ after multiple comparisons correction (all $p \geq 0.12$, all $|Z| \leq 2.06$; WSR).

An important caveat is that the RDS stimuli may not capture all units with some degree of tuning preference. As such, a conservative interpretation would be that group-level suppression during auditory attention is driven by units that do not exhibit strong tuning preferences. Additionally, both tuned and untuned populations contained units with significant evoked responses to the two TCs, although fractions of responsive units were higher in the tuned group (**Figure 6.D.c**). This shows that an absence of STRF tuning does not imply that units were not responsive to the task stimuli.

For the tuned group, does frequency preference determine degree of attentional modulation? We measured the best frequency (BF) of the excitatory field in each tuned STRF (**Figure 6.F**). Consistent with previous work showing that task demands shape frequency representation in AC (Atiani et al., 2009; Fritz et al., 2005, 2003; Yin et al., 2014), we found a

strong BF preference for a 1-octave band around the center frequency of the rewarded TC (**Figure 6.G**). Furthermore, distributions of BFs measured during the A-rule and V-rule were strikingly similar. This suggests that in our task, AC had shifted its frequency representation in a manner that was not rule-dependent. To test whether modulation by rule was dependent on tuning, we next divided units by their BF, as measured from the A-rule STRF, into groups near center frequency of $A_R$ (± 0.5 octaves), near $A_U$ or with a BF outside of either band. No difference between the groups was observed for responses to $A_R$ or $A_U$, although (Kruskal-Wallis non-parametric ANOVA, all $p>0.12$, all $H < 5.5$, FDR-adjusted *p*-values; **Figure 6 - source data 1.C,D**), suggesting that frequency tuning does not determine suppression or enhancement by attention in this task.

***Attention to sound increases encoding efficiency in deep-layer BS units.*** Previous work has established that FR changes do not necessarily imply changes in the amount of information spikes carry about sensory stimuli. For instance, optogenetic activation of inhibitory interneurons can reduce FRs in AC without changing information, suggesting increased encoding efficiency (Phillips and Hasenstaub, 2016). By contrast, locomotion reduces both FRs and information in AC (Bigelow et al., 2019). To determine whether reduced FRs evoked by decision stimuli were accompanied by changes in information or encoding efficiency, we used a PSTH-based neural pattern decoder to compare sound discrimination across attentional states (Foffani and Moxon, 2004; Hoglen et al., 2018; Malone et al., 2007). For each unit, the decoder generates a single-trial test PSTH and then compares these to two or more template PSTHs from distinct stimulus response conditions, generated sans test trial (**Figure 7.A**). The test trial is assigned to the template that is closest in *n*-dimensional Euclidean space, reflecting *n* PSTH bins. This is repeated for all trials, generating new templates for each classifier run. After all trials have been classified, a confusion matrix is generated. From this, we calculated accuracy of classification, mutual information (MI; bits) and encoding efficiency, a spike-rate normalized MI (bits/spike). As in previous analyses, a 0-300ms post-stimulus onset window was used in this method to restrict decoding to a predominantly sensory-driven component of the response. The binwidth for generating PSTHs was 30 ms (Hoglen et al., 2018). Only trials with correct responses (hits and CRs) and units with a minimum stimulus response FR of 1 Hz to both stimuli used in the decoder comparison were included.

We found that task rule could be decoded at greater than chance levels from responses to all four AV stimuli, and at all depth and NS/BS groups, showing that attentional state modulates decision stimulus PSTH responses throughout AC (**Figure 7 - figure supplement 1**; **Figure 7 - source data 1**). These comparisons suggest response modulation by task rule, but

do not address how information processing changes *across* the rules. To test this, we next used the decoder to compare accuracy in discriminating between responses to $A_R$* (rewarded in A-rule) and responses to $A_U$* (unrewarded in A-rule) bimodal stimuli across A-rule and V-rule conditions. This mimics the tone cloud discrimination required by the mice during the A-rule. In both rules, classification accuracy for the auditory decision stimuli ($A_R$*, $A_U$*) was higher than chance for all depth and BS/NS groups (see scatter plots in **Figure 7.B**; all p≤1.4e-05, all |z|≥4.2, one-way WSR vs chance [50%]; see **Figure 7 - source data 2.A** for stats). Sound classification accuracy ($A_R$*, $A_U$*) did not significantly differ *across* the A-rule and V-rule (**Figure 7.B**, $A_R$* *vs.* $A_U$* *comparison across rules:* all p≥0.17, all |z|≤1.39, see **Figure 7 - source data 3.A** for full stats; paired WSR on decoder accuracy in A-rule vs. V-rule by depth and NS/BS groups). Despite a reduction in activity levels during auditory attention, there was no loss in decoder accuracy, suggesting a possible change in encoding efficiency.

Through analysis of all decoder runs, we found that classifier accuracy and raw information were indeed correlated with FR (accuracy: r(3001) = 0.49, p = 2.3e-180; MI: r(3001) = 0.41, p = 1.5e-123; Pearson's correlation, all $A_R$* vs $A_U$* decoder runs). Thus, normalizing information by mean joint per-trial spike rate for the two responses in each decode (bits/spike) provides insight into the efficiency with which spikes are used to represent stimuli. We found that this encoding efficiency measure increased by ~20% during the A-rule for deep-layer BS units (**Figure 7.C**, $A_R$* *vs.* $A_U$* *comparison across rules:* Deep BS: p = 2.9e-04, Z = -4.06, paired WSR, FDR-adjusted *p*-value; med. FC: 1.19 [fold change: A-rule/V-rule]; V-rule: 0.15 ± 0.13, A-rule: 0.19 ± 0.19, mean bits/spk ± SDs, n = 233; all other groups p≥0.40, all |Z|≤1.47; see **Figure 7 - source data 4.A** for full stats). No other unit subpopulations showed significant changes. Note that for clarity, the above results are presented as the mean of decoder comparisons $A_RV_R$ vs. $A_UV_R$ and $A_RV_U$ vs. $A_UV_U$, thus collapsing visual stimulus identity. Analysis of these comparisons separately yields highly similar results (**Figure 7 - figure supplement 2; Figure 7 - source data 2, S9**), suggesting that visual stimulus identity does not contribute substantially to decoder accuracy or encoding efficiency at the level of group analysis.

***Receptive fields mapped during the inter-trial interval also show increased stimulus encoding efficiency.*** The analyses above revealed that auditory attention increased the per-spike encoding efficiency of task decision sounds. Does this effect of cross-modal attention switching generalize to encoding of sounds that were explicitly designed to be task-irrelevant? This helps determine whether attention observed here is specific to features of the auditory stream or broadly alters encoding of incoming auditory information. To address this, we tested whether information between STRFs derived from task-irrelevant ITI sounds and spiketrains

was modulated by attentional demands of the task. We restricted our analyses to only those units with STRFs passing the reliability criterion shown in **Figure 6.B.** To calculate STRF-spiketrain MI for each SU, we first calculated probability distributions of STRF-stimulus projection values for all stimulus time points (*P(x)*) and for those time points preceding a spike (*P(x|spike)*; **Figure 7.D**). Intuitively, these projection values reflect the similarity between a windowed stimulus segment at a given timepoint and the STRF. The divergence of the two projection distributions is captured in a spike-rate normalized MI measure (bits/spk; encoding efficiency), which describes the reliability with which spikes are determined by stimulus features of the STRF (**Figure 7.E**). No differences in encoding efficiency between conditions were observed in the superficial or middle BS/NS groups, or the deep NS group. Instead, consistent with our earlier findings for decision stimuli, encoding efficiency showed a significant A-rule increase in the deep BS subpopulation (**Figure 7.F**; Deep BS: p = 0.014, Z = -3.05, med. FC: 1.25, n = 50; paired WSR, FDR-adjusted *p*-value; fold change: A-rule/V-rule; mean bits/spk ± SDs; all other groups p≥0.24, all |Z|≤1.66; see **Figure 7 - source data 5** for full stats). This finding shows that during auditory attention, stimulus encoding is better described by a linear STRF filter and thus better tracks physical sound features. Furthermore, it suggests that increased encoding efficiency resulting from decreased spiking is a general effect of auditory attention in deep layer BS units, regardless of the context-based behavioral relevance or learned valence of the sounds.

***Information encoding efficiency changes are driven by suppressed units.*** The increase in A-rule encoding efficiency and decrease in average FRs in deep AC led us to further explore the relationship between activity level and information changes. Specifically, we tested whether group-level information efficiency changes are driven by SUs with suppressed responses, and how the minority of units with increased A-rule FRs perform in the decoder. We therefore examined classifier accuracy and encoding efficiency for target and distractor ($A_R$* vs $A_U$*) decoding separately for deep-layer BS units with increased and decreased FRs in the A-rule (**Figure 7 - figure supplement 3**). We found that units with increased FRs (39%; n = 96) exhibited a significant increase in A-rule decoding accuracy (**Figure 7 - figure supplement 3.C;** p = 0.0030, Z = -2.97, med. FC: 1.04, V-rule % correct: 66.4 ± 15.5, A-rule: 69.5 ± 15.5, n = 96; paired WSR; fold change: A-rule/V-rule; mean ± SDs), but no significant change in encoding efficiency (p = 0.84, Z = 0.2, V-rule bits/spk: 0.18 ± 0.15, A-rule: 0.18 ± 0.16). By contrast, units with suppressed FRs (60%; n = 146) showed no significant change in decoding accuracy (**Figure 7 - figure supplement 3.D**; p = 0.44, Z = 0.77, V-rule: 67.32 +/- 15.25, A-rule: 66.53 +/- 14.61), but a 44% increase in encoding efficiency (p = 1.8e-07, Z = -5.22, med. FC: 1.44, V-rule:

0.13 +/- 0.12, A-rule: 0.18 +/- 0.19; paired WSR). These results suggest that the minority of units that increase FR in the A-rule perform marginally better at decoding the auditory stimulus, and that the units that decrease FR drive the shift in encoding efficiency.

***Attention-related firing rate changes predict correct task performance.*** To ensure that mice were adequately engaged and attentive in the task, the analyses described above excluded any trials in which the incorrect behavioral response was made. However, an examination of these error trials, which may correlate with lapses in attention, could provide insight into the moment-to-moment behavioral relevance of the attentional effects described above. We have shown that attention to sound is marked by a net suppression of pre-stimulus and evoked FRs. We hypothesized that, if this attentional modulation is behaviorally meaningful, FRs preceding A-rule error trials may be more similar to sound-unattended V-rule trials than to A-rule correct trials. We addressed this possibility by comparing pre-stimulus FRs in error versus correct trials (300 ms prior to stimulus onset; **Figure 8.B**). Because misses were uncommon (**Figure 8.A**), we restricted our analysis to the comparison of FA and CR trials to allow for adequate sampling of each trial outcome. We included only behavior sessions with at least 10 FA and CR trials (A-rule and V-rule trials considered separately). This decreased unit sample sizes (n = 234, 58 across all depth groups for BS, NS; min. group size = 9, 2 for BS, NS). Given the small sample of NS units and the likelihood of insufficient power, NS units were not included in this analysis. When considering BS units *with increased FRs in the A-rule*, we found no significant group-level difference between A-rule FA and CR trials at any cortical depth (**Figure 8C; Figure 8 - source data 1.A** for full stats; all p≥0.29, all |Z|≤1.43, paired WSR, FDR-adjusted *p*-values). However, deep cortical BS units *with A-rule suppression* showed significantly higher pre-stimulus FRs prior to A-rule FA trials than CR trials (Deep BS [n = 98]: mean FR difference between pre-stim FA and CR trials = 0.35 Hz, p = 0.0098, Z = -3.15; paired WSR; other depth groups: p≥0.28, |Z|≤1.48; all *p*-values FDR-adjusted). This is unlikely to reflect a motor effect of higher FR before a lick, as it was specific to the A-rule: pre-stimulus FRs in A-rule-suppressed or A-rule-enhanced units did not differ between FA and CR trials in the V-rule (**Figure 8.C.c; Figure 8 - source data 1.B**; paired WSR: all p≥0.68, all |z|≤1.59, FDR-adjusted *p*-values). Together, these findings suggest that FR reductions typical of modality-selective attention directly relate to behavioral outcomes.

# Discussion

In the present study, we recorded SU activity across AC layers in mice performing an audiovisual rule-switching task. We compared responses evoked by identical stimuli under

conditions of auditory or visual modality-selective attention. Attention to sound shifted AC stimulus representation by decreasing activity of untuned units and increasing encoding efficiency in the deep cortical laminae. Pre-stimulus activity was also reduced by auditory attention, which accounted for changes in stimulus-evoked responses. The effects of attention extended beyond the decision stimuli required to complete the task; responses to task-irrelevant receptive field mapping stimuli exhibited similar reductions in evoked activity and increases in encoding efficiency, suggesting that attention to sound induces a stimulus-general shift in processing. This attentional shift was behaviorally meaningful, with error trials in the A-rule predicted by higher firing rates in the set of units that is suppressed under auditory attention. Taken together, these results show that attending to sound results in a general suppression of ongoing activity in AC, while retaining activity critical for sensory representation.

Attentional highlighting of behaviorally-relevant signals may employ multiple mechanisms, including response enhancement or noise suppression. Feature selective attention studies have shown that firing rates for neurons tuned to attended features are often increased, thereby increasing the reliability of the sensory cortical readout (Desimone and Duncan, 1995; Moran and Desimone, 1985; Reynolds and Chelazzi, 2004). Another mechanism which may act in tandem with response enhancement is the reduction of noise to improve encoding reliability. Noise reduction may act through decreased rates in pre-stimulus baseline activity (Buran et al., 2014), reduced variance in single neuron rates (Mitchell et al., 2007) or decreased correlations of noise across the population (Cohen and Maunsell, 2009; Downer et al., 2015). In the present study, we found no evidence for increased signal-to-noise ratio in the firing rate signal, as shown by the roughly equal stimulus response magnitudes across rules when normalizing for pre-stimulus rate. However, the timing of activity in AC is known to carry substantial information (Hoglen et al., 2018; Krishna and Semple, 2000; Malone et al., 2007), which would not be captured by coarse rate estimations. By accounting for fine scale temporal patterns with a PSTH-based pattern classifier and analysis of stimulus-STRF selectivity, we show that decreased ongoing activity and a concomitant increase in encoding efficiency at the group level provides an additional mechanism for attentional noise reduction, perhaps refining the stimulus-encoding portion of the neural signal for readout in downstream brain areas.

Previous studies of behavioral state-dependent state changes in auditory processing have typically compared task-engaged and passive sound processing. While this paradigm does not specifically isolate the effects of attention due to confounds of arousal, attention, reward expectation, and motor activity (Saderi et al., 2021), it has provided valuable insight into the dependence of sensory processing on task-engaged behavioral states. Consistent with our

findings, this work has shown that AC stimulus-evoked spiking responses are predominantly suppressed during self-initiated task engagement when compared to passive listening (Bagur et al., 2018; Carcea et al., 2017; Kuchibhotla et al., 2017; Otazu et al., 2009). Activity levels preceding a stimulus may also decrease (Buran et al., 2014; Carcea et al., 2017), although some studies in AC do not show this effect (Otazu et al., 2009). Reductions of pre-stimulus activity during task engagement have also been observed in rat gustatory cortex (Yoshida and Katz, 2011) and monkey visual cortex (Bisley and Goldberg, 2003; Cox et al., 2019; Herrington and Assad, 2010; Sato and Schall, 2001).

Neuronal stimulus preferences relative to a target have been shown to determine the degree of attentional modulation, such that stimulus-evoked responses for attended features are generally enhanced but can also be suppressed for features outside of the receptive field (Reynolds and Chelazzi, 2004). Here, we find that frequency preferences of units with STRF tuning do not appear to determine suppression or enhancement within the task, but critically we also find that the bulk of units with STRF tuning exhibit a preference for frequencies near the rewarded TC (**Figure 6G**). This is consistent with a body of work from Shamma, Fritz and colleagues showing that engagement in an auditory discrimination task rapidly shifts AC receptive fields to enhance frequency representation of behaviorally-relevant stimuli (Atiani et al., 2009; Fritz et al., 2005, 2003; Yin et al., 2014). In our task, mice were trained for multiple months prior to physiological recordings, and the tone cloud frequencies of rewarded and unrewarded stimuli were held consistent for each animal. As such, spectral representation in the AC of our highly trained mice is biased toward task-relevant stimuli. Speculatively, it is possible that tuning-dependent attentional modulation may occur in earlier stages of task acquisition, but that the substantial reconfiguration of sound processing tailored to the task alters its expression after training. The distribution of preferred frequencies also does not shift between auditory and visual rules, suggesting that attending to visual stimuli does not place plasticity-inducing demands on AC frequency representation. Instead, we find that units without STRF tuning drive the reduction in neural activity during auditory attention. An important caveat is that our STRF-based approach is only one way to determine AC tuning and other stimulus and analysis methods may reveal additional tuning preferences. Nevertheless, we believe that this method provides a useful classification for degree of tuning. This result is also consistent with our information theoretic analyses in that both suggest that attention to sound may selectively remove spikes that are minimally sound-driven.

As in previous studies, attention-related modulation was not uniformly expressed across cortical depths and neuron types. Changes in both FR and encoding efficiency were most

prominent in deep layer neurons. These findings extend several previous studies reporting larger effects of attention in infragranular LFP and multi-unit activity (O'Connell et al., 2014; Zempeltzi et al., 2020). These physiological outcomes are consistent with anatomical work suggesting top-down modulatory signals arrive primarily in the supragranular and infragranular layers (Felleman and Van Essen, 1991). As the main cortical output layer, information shifts in the infragranular population would differentially influence subcortical sites and other cortical regions (Salin and Bullier, 1995). One important caveat is that superficial AC is known to have lower spontaneous and evoked FRs than deeper cortex (e.g. **Figure 4C**; Christianson et al., 2011; Sakata and Harris, 2009), which may have made it more difficult for us to observe statistically significant attention-related effects. Furthermore, although we tried to minimize neural tissue damage through technical considerations such as using a slow probe insertion speed (Fiáth et al., 2019), the superficial layers likely sustain the greatest level of damage when the probe is inserted to span the full cortical depth. Despite these factors, we were able to isolate a reasonably large sample size of responsive neurons in superficial cortex from successful behavior sessions (n = 119 units, of which 57% were stimulus-responsive). Nevertheless, we cannot rule out whether the absence of observed attentional modulation at superficial depths may have been due to experimental limitations such as the comparatively small sample size. Future work employing imaging techniques to target superficial neurons may help resolve this.

Previous studies have reported larger effects of task engagement or attention in inhibitory interneurons (Kuchibhotla et al., 2017; Mitchell et al., 2007). As such, attention-related reduction of activity could be sustained by inhibitory network drive. Our approach of dividing activity into broad- and narrow-spiking does not suggest a general increase in NS activity during auditory attention. However, we observed heterogenous types of modulation; in many units, NS activity decreased during auditory attention, but in a smaller group, there was a significant increase. An important caveat is that the BS/NS distinction is an imperfect approximation of excitatory/inhibitory activity, with many inhibitory cell types presenting a BS waveform phenotype (e.g. somatostatin-positive interneurons [Li et al., 2015]). An alternative mechanism is that excitatory drive is decreased during auditory attention. These two proposed mechanisms – increased inhibitory tone and decreased excitatory drive - are not mutually exclusive.

Our findings suggest that attentional selection is achieved by removal of a noise background on which sound stimulus-encoding activity sits. This is in line with an influential theory of cortical attention that posits that spontaneous activity fluctuations partly reflect internal processes such as mental imagery or memory recall, in contrast with activity that arises from

external sensory stimulation (Harris and Thiele, 2011). In this model, attention suppresses internally-generated spontaneous activity to favor the processing of behaviorally-relevant external stimulation. The work presented here offers several pieces of evidence in favor of this theory. Auditory attention suppresses activity in untuned units, affecting both pre-stimulus and stimulus-evoked activity. This activity reduction does not alter stimulus-spike train decoding accuracy, but instead increases stimulus encoding efficiency and preserves stimulus representation.

In summary, we demonstrate a novel connection between attention-induced shifts in activity levels and stimulus encoding in early sensory cortex, which are directly related to behavioral outcomes. Previous research suggests such outcomes reflect top-down control by executive networks comprising frontal, parietal, thalamic, and striatal areas (Cools et al., 2004; Crone et al., 2006; Licata et al., 2017; Rikhye et al., 2018; Rougier et al., 2005; Toth and Assad, 2002; Wimmer et al., 2015). These networks may act as a context-dependent switch, routing attentional modulatory feedback to the sensory systems. In the present study, we provide evidence that such modulation specifically suppresses stimulus-irrelevant spiking, thus enhancing encoding efficiency in deep AC neurons.

# Methods

**Animals.** All experiments were approved by the Institutional Animal Care and Use Committee at the University of California, San Francisco. Twenty-seven C57BL/6 background male mice were surgically implanted with a headpost and began behavioral training, of which 10 completed the training and successfully performed the task during physiology recording sessions. All mice began the experiment between ages P56 and P84. Mice used in this report expressed optogenetic effectors in various subsets of interneurons, which we intended to use for optogenetic identification of cells (Lima et al., 2009; analysis not included here). These mice were generated by crossing an interneuron subpopulation-specific Cre driver line (PV-Cre JAX Stock Nr. 012358; Sst-Cre: JAX Stock Nr. 013044) with either the Ai32 strain (JAX Stock Nr. 012569), expressing Cre-dependent eYFP-tagged channelrhodopsin-2, or the Ai40 strain (JAX Stock Nr. 021188), expressing Cre-dependent eGFP-tagged archaerhodopsin-3. Of the 10 behavior mice included in this report, 6 were Ai32/Sst-Cre, 3 were Ai32/PV-Cre and one was Ai40/Sst-Cre. In most experiments (n = 21 recordings), brief, low-level optogenetic pulses during the inter-trial interval of the task were used to identify opsin-expressing neurons (<0.3 mW light; 5 light pulses of 10 ms duration, every ~1.5 min); these analyses are outside of the

scope of this report. The optogenetic stimulation protocol was consistent through A- and V-rules of the task. Unit stimulus response FRs and behavioral response error rates were not statistically different between trials immediately after optogenetic pulses and stimulus-matched trials preceding the pulses.

All mice were housed in groups of 2-5 for the duration of the behavioral training until the craniotomy. Post-craniotomy and during physiology recordings mice were housed singly (up to 6 days) to protect the surgical site. Mice were kept in a 12 hr/12 hr reversed dark/light cycle. All training occurred during the dark period, when mice showed increased activity and behavioral task performance (Roedel et al., 2006).

**Audiovisual rule-switching behavior task**. Adult mice (>P56) were trained on an audiovisual (AV) go/no-go rule-switching behavior task. In this task, mice were positioned on a floating spherical treadmill in front of a monitor and a speaker, and an optical computer mouse recorded treadmill movement. Mice licked to receive a reward depending on auditory, visual or AV stimulus presentation ("decision" stimuli, either "target" or "distractor"), but the modality predictive of the reward changed partway through the behavioral session. Each session would start with a unimodal go/no-go block, in which a series of auditory ($A_R$, $A_U$; 17kHz or 8 kHz tone clouds [TC]) or visual ($V_R$, $V_U$; upward or rightward moving gratings) stimuli was presented. After stimulus presentation, mice signaled choice by either licking a spout in front of the mouth or withholding licking. Licking at the target unimodal stimulus would trigger a water reward, while licking at the distractor would trigger a short dark timeout. After a fixed number of unimodal trials, the stimuli would become AV, but the rule for which stimulus predicted reward would carry over from the unimodal block. All four stimulus combinations ($A_RV_R$, $A_RV_U$, $A_UV_R$, $A_UV_U$) would be presented in the AV block, such that two AV combinations would be target stimuli and two would be distractor. Then, after completing a fixed number of trials in the AV block, the task using the rule of the opposite modality would begin; a unimodal block with the other modality would start, followed by a second AV block using the rule from the preceding unimodal block. For any mouse, the stimuli predictive of the reward in each rule was kept constant across days and training sessions (e.g., a 17 kHz TC would always predict a reward in the A-rule, and a rightward grating would always predict a reward in the V-rule). In most cases, blocks proceed without interruption, although in some cases, the task is paused and resumed momentarily if the mouse fails to correctly transition between rules.

The task was self-paced using a virtual foraging approach, in which mouse locomotion (measured through treadmill rotation) would cause a track of randomly placed dots on the screen to move down. After a randomly varied virtual distance, a decision stimulus would be presented, at which point the mouse would lick or withhold licking to signal choice. For receptive field mapping during physiology experiments, a random double-sweep (RDS) stimulus was presented in between decision stimuli, during the inter-trial track portion. Stimuli are detailed below.

**Behavior training and apparatus.** Prior to any training, mice were surgically implanted with a stainless steel headplate, used both for head fixation during the task, and for physiology recordings after the task was learned (methods described below). Three days post-implant, mice began a water restriction protocol based on previously published guidelines (Guo et al., 2014). Throughout the course of training, mice received a minimum water amount of 25 mL/kg/day, based on weight at time of surgical implant. After recovery from surgery, mice were given ~7 days to adjust to water restriction. Then, mice were head fixed and habituated to the floating treadmill for 15-30 min daily sessions with no stimulus presentation for 2-3 days. After mice appeared comfortable on the treadmill, a phased behavioral task training regimen began. Mice were trained once daily for ~6 days per week. On day 1, mice were introduced to an auditory-only (A-only) stimulus training version of the task in which $A_R$ ("target"/"rewarded") or $A_U$ ("distractor"/"unrewarded") stimuli were presented, and a reward would be automatically administered shortly after the onset of $A_R$. Next, the mice were put on an operant version of the A-only task, which required licking any time after the onset of $A_R$ to receive a reward and withholding of licking during $A_U$ to avoid a dark timeout punishment. Mice achieved proficiency, defined as two or more consecutive days of sensitivity index d' > 1.5 (see *Data analysis* for calculation), on the A-only task after 11.0 ± 4.7 d after start of training (median ± SD, n = 10 successful mice). Then, a similar training structure was repeated for the visual task: V-only stimulus training with automatic rewards for $V_R$, but not $V_U$, followed by an operant version of the visual task requiring licks for rewards (median time to proficiency: 26.0 ± 7.2 d after start). After learning the tasks for each modality separately, mice were introduced to an auditory-AV (A-AV) version, in which the rule from the auditory stimulus carried over to the AV block. This was intermixed with training days on a visual-AV (V-AV) version of the task. Number of training days on A-AV or V-AV were decided based on prior performance, with extra training given as needed. Mice were considered proficient at this stage after performing with d'>1.5 on each rule (A-AV; V-AV) on two consecutive days (median time to proficiency: 40.0 ± 15.8 d after start).

Finally, the full rule-switching task was introduced (**Figure 1.D**), generally alternating between days of V-rule-first and the A-rule-first task sequences but allocating more training days to task orders as needed. Because physiology recordings were acute and strictly limited to 6 days after craniotomy, we set a greater threshold for expert-level performance on the full task before advancing to physiology: three consecutive days of d'>2.5 (median time to expertise: 90.5 ± 31.8 d). Care was taken to train each mouse at a roughly consistent time of day (no more than ~1-2 hrs day-to-day variation). During expert-level task performance, mice typically completed 260-300 trials in a daily session (30 A-only; 100 to 120 A-AV; 30 V-only; 100 to 120 V-AV).

The behavior training setup was controlled by two computers: a behavior monitoring and reward control PC (OptiPlex 7040 MT, Dell) and a dedicated stimulus presentation machine running Mac OS X (Mac Mini, Apple). Stimulus presentation was controlled with MATLAB using custom software (Morrill et al 2022), and inter-machine communication used the ZeroMQ protocol. Auditory and visual stimuli were generated and presented using the Psychophysics Toolbox Version 3 (Kleiner et al., 2007). Water rewards were administered using a programmable syringe pump (NE-500, New Era Pump Systems, Farmingdale, NY), positioned outside of the sound-attenuating recording chamber. Early in training, water reward volume was set at 0.01 mL per correct response, but over training the reward volume was gradually decreased to 0.006 mL to achieve greater trial counts. Licking events were recorded using a custom photobeam-based lickometer circuit based on plans provided by Evan Remington (Xiaoqin Wang Lab, Johns Hopkins University). Licks were registered when an IR photobeam positioned in front of the lick tube was broken, queried at a sample rate of 100 Hz by an Arduino Uno microcontroller (Arduino, LLC).

***In vivo* awake recordings during behavior.** Animals in this experiment underwent two surgeries: first, before training a surgery to implant a custom steel headplate over the temporal skull using dental cement was conducted. The animal was anesthetized using isoflurane and a headplate was positioned over AC, ~2.5 mm posterior to bregma and under the squamosal ridge, to allow for physiology recordings after achieving task expertise. When mice completed the training regimen outlined above, a craniotomy surgery was performed. The animal was again anesthetized using isoflurane and an elliptical opening (0.75 mm wide x 1.5 mm long) was made in the skull over AC using a dental drill. This opening was promptly covered with silicone elastomer (Kwik-Cast, World Precision Instruments), and the animal was allowed to recover overnight. The following day, the animal was affixed by its headplate over the treadmill inside of a sound-attenuating recording chamber, the silicone plug over the craniotomy was removed and

the craniotomy was flushed with saline. A silver-chloride ground wire was placed into the craniotomy well at a safe distance from the exposed brain. A 64-channel linear probe (20 μm site spacing; Cambridge Neurotech, Cambridge, UK) was slowly inserted in the brain using a motorized microdrive (FHC, Bowdoin, ME) at an approximate rate of ~1 μm / s (Fiáth et al., 2019). After reaching the desired depth, the brain was allowed to settle for 10 min, after which the water spout, lickometer, visual stimulus delivery monitor and speaker were positioned in front of the mouse, and the behavior session commenced. Behavior sessions were sometimes stopped early and restarted due to poor performance. In approximately half of behavior-physiology sessions (13 of 23 successful recordings), the task was stopped due to low performance after the rule transition and restarted at the beginning (unimodal block) of the second rule. To control for possible effects of task order, attempts were made to counterbalance recordings from A-rule first (15) and V-rule first (8) behavior sessions.

After completion of the behavior task, the water spout and lickometer were removed, and a series of auditory and/or visual passive experiments were conducted in order to characterize the response properties of the recording site. All stimuli were presented with the auditory and visual stimulation apparatus described above. Following completion of these experiments, the probe was slowly removed, and the brain was covered with a thin layer of freshly mixed 2.5% agarose in saline, followed by a layer of silicone elastomer. The animal was returned to its home cage, and the following day the physiological recording process was repeated. Recordings were made for up to 6 days after the craniotomy. The neural signal acquisition system consisted of an Intan RHD2000 recording board and an RHD2164 amplifier (Intan Technologies), sampling at 30 kHz.

**Auditory and visual stimuli**. In-task auditory decision stimuli were 1 s TCs, consisting of 50 ms tone pips overlapping by 25 ms, with frequencies in a 1 octave band around either 17 kHz or 8 kHz. TCs were frozen for the duration of the task, so that each mouse always heard the same pip sequences, allowing for direct comparisons of sound-evoked neural responses across rules without concern that stimulus peculiarities may be driving observed differences. TCs were presented at 60 dB SPL. Visual decision stimuli consisted of a circular moving grating stimulus (33° diameter subtended visual space), which appeared at the center of the screen for 1 s (coincident with TC stimulus during bimodal presentation). Gratings moved either upward or rightward with a 4 Hz temporal frequency, 0.09 cycles/degree spatial frequency at 50% contrast. In between decision stimulus presentations, a random double-sweep (RDS) stimulus was presented for receptive field mapping (Bigelow et al., 2022; Gourévitch et al., 2015). The RDS comprised two uncorrelated random sweeps that varied continuously and smoothly between 4

and 64 kHz, with a maximum sweep modulation frequency of 20 Hz. RDS stimuli were presented at 50 dB SPL.

After the behavior task, passive auditory search stimuli (pure tones, click trains) were presented to characterize response properties of the electrode channel. Click trains consisted of broadband 5 ms white noise pulses, presented at 20 Hz for 500 ms duration. Pure tone stimuli consisted of 100 ms tones of varied frequencies (4 – 64 kHz, 0.2 octave spacing) and sound attenuation levels (30 – 60 dB in 5 dB linear steps), with an interstimulus interval of 500 ms.

Auditory stimuli were presented from a free-field electrostatic speaker (ES1, Tucker-Davis Technologies) driven by an external soundcard (Quad-Capture or Octa-Capture, Roland) sampling at 192 kHz. Sound levels were calibrated using a Brüel & Kjær model 2209 meter and a model 4939 microphone. Visual stimuli were presented on a 19-inch LCD monitor with a 60Hz refresh rate (Asus VW199), positioned 25 cm in front of the mouse and centered horizontally and vertically on the eyes of the mouse. Monitor luminance was calibrated to 25 cd/m$^2$ for a gray screen, measured at approximate eye level for the mouse.

**Data analysis.** *Behavioral performance.* Task performance was evaluated by calculation of the *d'* sensitivity index:

$$d' = Z(H) - Z(F)$$

where *H* is hit rate and *F* is false alarm rate, and *Z* is the inverse normal transform. Because this transform is undefined for values of 0 or 1 and hit rates of 1 commonly occurred in this study, we employed the log-linear transformation, a standard method for correction of extreme proportions, for all calculations of *d'* (Hautus, 1995). In this correction, a value of 0.5 is added to all elements of the 2 x 2 contingency table that defines performance such that:

$$H = (hits + 0.5)/(hits + misses + 1)$$

$$F = (FA + 0.5)/(FA + CR + 1)$$

where *FA* is the false alarm count and *CR* is the correct reject count. To ensure that mice properly transitioned between task rules, *d'* values were calculated separately for responses in the A-rule and the V-rule. Behavioral sessions during physiological recording with *d'* < 1.5 in either rule were excluded from analyses, as were any sessions with a false alarm rate > 0.5 to

stimuli with conflicting reward valances across rules: $A_UV_R$ in A-rule or $A_RV_U$ in V-rule (n = 23 successful sessions, n = 10 mice; 1 session excluded due to recording artifact, see below).

*Spike sorting and unit stability evaluation.* Spikes were assigned to unit clusters using KiloSort2 (KS2; Pachitariu et al., 2016). Clusters were first evaluated for isolation quality through the automated KS2 unit classification algorithm and then with a custom MATLAB interface. In this second step, clusters with non-neuronal waveforms or 2 ms refractory period violations >0.5% were removed from analysis (Laboy-Juárez et al., 2019; Sukiban et al., 2019). To evaluate stability, activity for each unit was plotted for the recording duration as a raster and binned spike counts (2 min bins) and manually examined for periods with a substantial drop off in FR (periods flagged for instability: 88 ± 10 % [mean ± SD] decrease in FR from median activity level). Flagged unstable periods were marked and removed from analysis (101/ 742 SUs with flagged durations >10% of recording time). One session meeting behavior performance criteria was excluded due to a high degree of electrical noise contamination.

*Classification of units by depth and waveform shape.* Probes with electrode spans of 1260 µm were used, allowing for channels below and above AC. During recording, the probe was lowered to a point where several channels showed a prominent drop in field potential amplitude and spiking activity, indicating penetration into the white matter (Land et al., 2013). After behavior sessions, a set of auditory and visual stimulation protocols was used to map response properties of each electrode site, and multi-unit activity (MUA) responses were analyzed. Here, we define MUA as threshold crossings of 4.5 SD above a moving window threshold applied to each channel. Analysis of MUA was restricted to site characterization and is not included in the main results. We analyzed each tone or click PSTH for reliable responses, which we defined as trial-to-trial similarity of $p<0.01$ (Escabí et al., 2014). We designated the deepest channel with a reliable MUA sound response of any magnitude as the deep cortex-white matter border. Limited somatic spiking in the top layer of cortex prevented the use of MUA as a reliable marker for the superficial cortex-pia border (Senzai et al., 2019), so we instead relied on a LFP-based measure. To define the top border of cortex, the maximum spontaneous LFP (1-300 Hz) amplitude of a 10 s snippet from each channel was plotted, and the channel at which LFP amplitude dropped off to the approximate probe-wise noise floor (i.e., minimum LFP amplitude) was considered the top channel in cortex (**Figure 3.B.c**). These measures were confirmed histologically through Di-I probe marking experiments with a separate group of untrained mice; histology methods described below and elsewhere (Morrill and Hasenstaub, 2018)**.** Marking the top and bottom cortical borders generated a span of channels putatively within AC. This span

was used to divide channels into superficial, middle, and deep groups, based on measurements of the fraction of cortex attributed to supragranular (layers 1-3), granular (layer 4) and infragranular (layers 5-6) in the mouse AC (Allen Institute Mouse Brain Atlas; https://mouse.brain-map.org/). SUs were assigned the fractional depth of the channel on which the largest magnitude waveform was recorded.

Clusters were also classified into broad-spiking (BS; putatively excitatory) and narrow-spiking (NS; putatively fast-spiking inhibitory) units on the basis of the bimodal distribution of waveform peak-trough durations (**Figure 3.D;** NS/BS transition boundary = 0.6 ms). From sessions with successful behavior, we recorded 742 SUs from all cortical depths, comprising 17.5% (130) NS units and 82.5% (612) BS units.

*Firing rate analysis and trial filters.* To compare FR responses to stimuli across task rules and to the receptive field mapping stimulus, we measured FR in the first 300 ms post-stimulus onset. Only units with nonzero FRs in both rules were included. To ensure that measurements were capturing periods of task engagement, all trials with incorrect responses (misses and FAs) were excluded from all decision-stimulus analyses, with the exception of those shown in **Figure 8**. We also excluded trials with recorded licks earlier than the 300 ms post-stimulus onset, or in the 500 ms pre-stimulus onset. Given these filters, analyses were restricted to units present in the recording during at least 10 trials (correct behavioral choice and without “early licks”) for each stimulus type.

*PSTH-based Euclidean distance decoding.* A peristimulus time histogram (PSTH)-based decoder was used to compute the mutual information between spike trains and stimulus identity (**Figure 7.A**; (Foffani and Moxon, 2004; Hoglen et al., 2018; Malone et al., 2007). In this method, two or more responses are compared by generating template PSTHs by removing one test trial. This test trial response is also binned into a single-trial PSTH, and then classified as belonging to the nearest template in *n*-dimensional Euclidean space, where *n* is the number of PSTH bins. More formally, the nearest template is that which minimizes the Euclidean norm between test and template vectors (PSTHs). This process is then repeated for all trials comprising the template PSTHs. Decoding accuracy is the percentage of trial responses that are correctly assigned to the stimuli that elicited them. Mutual information (MI) is calculated from a confusion matrix of classifications as follows:

$$MI = \sum_i \sum_j P(X_i Y_j) log_2(\frac{P(X_i Y_j)}{P(X_i) * P(Y_j)})$$

where $X$ is the decoder prediction, $Y$ is the actual, $P(X_iY_j)$ represents the value of the ($i$, $j$) element of the confusion matrix, and $P(X_i)$ and $P(Y_j)$ are sums on the marginals. This yields a value of mutual information in bits. To measure encoding efficiency (bits per spike), we normalized mutual information by the joint mean spikes per trial of the responses submitted to the decoder (Bigelow et al., 2019; Buracas et al., 1998; Zador, 1998).

For consistency with FR analyses, a time window of 0 to 300 ms, where stimulus onset is 0, was chosen for decoding analysis. A PSTH binwidth of 30 ms was chosen based on optimal binwidth calculations for mouse AC using the same decoding method (Hoglen et al., 2018). To filter out units with low responsiveness to any of the stimuli in a given decoding analysis, we required a minimum FR of 1 Hz during the 0-300 ms window in both stimulus conditions. As such, unit sets may differ between each decoding analysis due to units that were responsive to one set of stimuli but unresponsive to others.

*Spectrotemporal receptive field (STRF) analysis.* To test whether task rule modulates auditory receptive fields, we presented a random-double sweep stimulus (RDS, described in *Auditory and visual stimuli*) in between trials for durations of ~1 to 15 s, depending on rate of task progression. Different randomly generated RDS segments were presented in each inter-trial interval, and STRFs were generated separately for each rule. Because total RDS duration varied between the A-rule and the V-rule in a single session, we equated presentation time across rules by truncating the segments of the rule with greater RDS time (presentation time in each rule: 6.8 ± 2.6 min [mean ± SD]; n = 23 sessions). This ensured that different stimulus presentation times did not bias STRF estimation. The first 200 ms of RDS response was dropped from all STRF analyses to minimize bias from strong onset transients. SU activity during these short RDS segments was used to generate STRFs for each segment using standard reverse correlation techniques (Aertsen and Johannesma, 1981; de Boer, 1968; Gourévitch et al., 2015). In brief, the spike-triggered average (STA) was calculated by summing all stimulus segments that preceded spikes using a window of 200 ms before and 50 ms after each spike. The choice of 200 ms prior to each spike reflects the upper limit of temporal integration times of auditory cortical neurons (Atencio and Schreiner, 2013), and the 50 ms post-spike time was included to estimate acausal values, i.e., those that would be expected by chance given the stimulus and spike train statistics (Gourévitch et al., 2015). STRFs were

transformed into units of FR (Hz) using standard methods discussed elsewhere (Rutkowski et al., 2002). Units with poorly-defined STRFs were filtered out using a trial-to-trial correlation metric (Escabí et al., 2014): STRF segments were randomly divided into two halves, re-averaged separately, and a correlation value was calculated for the two STRFs. This process was then repeated 1000 times, and the mean of correlations defined the reliability value for each STRF. We compared the mean observed STRF reliability to a null distribution of reliabilities, generated by repeating the procedure on null STRFs made from circularly-shuffled spike trains (preserving spike count and interspike interval but breaking the timing relationship between spikes and stimulus). A *p*-value was calculated as the fraction of the null STRF reliabilities greater than the mean observed STRF reliability, and STRFs with $p<0.05$ in either rule were included in subsequent analyses. Any STRFs from units with greater than 10% of recording duration marked as unstable were removed from analysis.

Mutual information between a spiketrain and a STRF was measured as the divergence of two distributions: one reflecting the similarity of the windowed stimulus segments (RDS) preceding a spike and the STRF, and the other reflecting the similarity of all possible windowed stimulus segments and the STRF, regardless of whether a spike occurred (**Figure 7.D**; (Atencio et al., 2008; Atencio and Schreiner, 2012; Escab□and Schreiner, 2002). Stimulus-STRF similarity was defined as the inner product of the STRF and the stimulus segment of equivalent dimensions, with higher values reflecting closer matches between the STRF and stimulus. The distribution *P(z|spike)* was generated from $z = s \cdot STRF$, where *s* represents all RDS stimulus segments that preceded a spike. Then the distribution *P(z)* was made from similarity calculations of all possible windowed RDS segments and the STRF. The mean $\mu$ and the standard deviation (SD) $\sigma$ of *P(z)* were calculated, and the distributions were transformed into units of SD: $x = (z - \mu)/\sigma$, yielding distributions of *P(x|spike)* and *P(x)* expressed in units of SD.

Using the distributions described above, a spike count-normalized measure of mutual information between the calculated STRF and the spike train can be calculated as:

$$MI = \sum P(x|spike) log_2(\frac{P(x|spike)}{P(x)})$$

We used this value to compare how well STRFs from A-rule and V-rule ITIs predict a spiketrain, and thus whether activity in each attentional condition is well described by this canonical filter model.

**Statistics.** All statistical calculations were performed in MATLAB r2019a and its Statistics and Machine Learning Toolbox, V11.5. For group comparisons of SU responses across task rules, paired Wilcoxon signed-rank (WSR) tests were used, unless otherwise noted. Because tests were performed separately on each depth and spike waveform subpopulation, the Benjamini-Hochberg false discovery rate (FDR) procedure was used to correct for multiple comparisons, typically across n = 6 comparisons (3 depth groups, 2 spike waveform groups; Benjamini and Hochberg, 1995). This method relies on controlling the Type I error rate (here, q = 0.05), providing increased power over typical family-wise error rate controls. To determine if individual SUs were significantly modulated by rule, an unpaired Student's t-test on FR was used with a threshold of $p<0.01$. Descriptive statistics reported in text are mean ± standard deviation (SD), unless otherwise noted. Fractional change values between task rules are reported as the median of the A-rule/V-rule. All other statistical tests are described in Results. Sample sizes (*n*) are indicated for each comparison in Results or Supplemental tables.

**Histological verification of depth measurement.** To test the accuracy of our depth estimation method based on physiological responses (**Figure 3**), we presented the pure tone search stimuli described above to a separate set of untrained control mice while performing extracellular recordings (n = 11 recordings from 4 mice; Ai32/Sst-Cre). Before insertion, the probe was painted with the fluorescent lipophilic dye Di-I (DiCarlo et al., 1996; Morrill and Hasenstaub, 2018). The depth measurement procedure based on physiological signals was carried out as described above, and then probe tracks from each recording were visualized as described previously (Morrill and Hasenstaub, 2018). Briefly, after recordings, the animal was euthanized, and the brain was removed and placed into a solution of 4% PFA in PBS (0.1 m, pH 7.4) for 12 h, followed by 30% sucrose in PBS solution for several days. The brain was then frozen and sliced using a sliding microtome (SM2000R, Leica Biosystems) and slices were imaged with a fluorescence microscope (BZ-X810, Keyence). Di-I probe markings showing cortical depth were consistent with physiological activity-based depth measurements described above (**Figure 3.B-C**).

# Figure list



# Source data table list

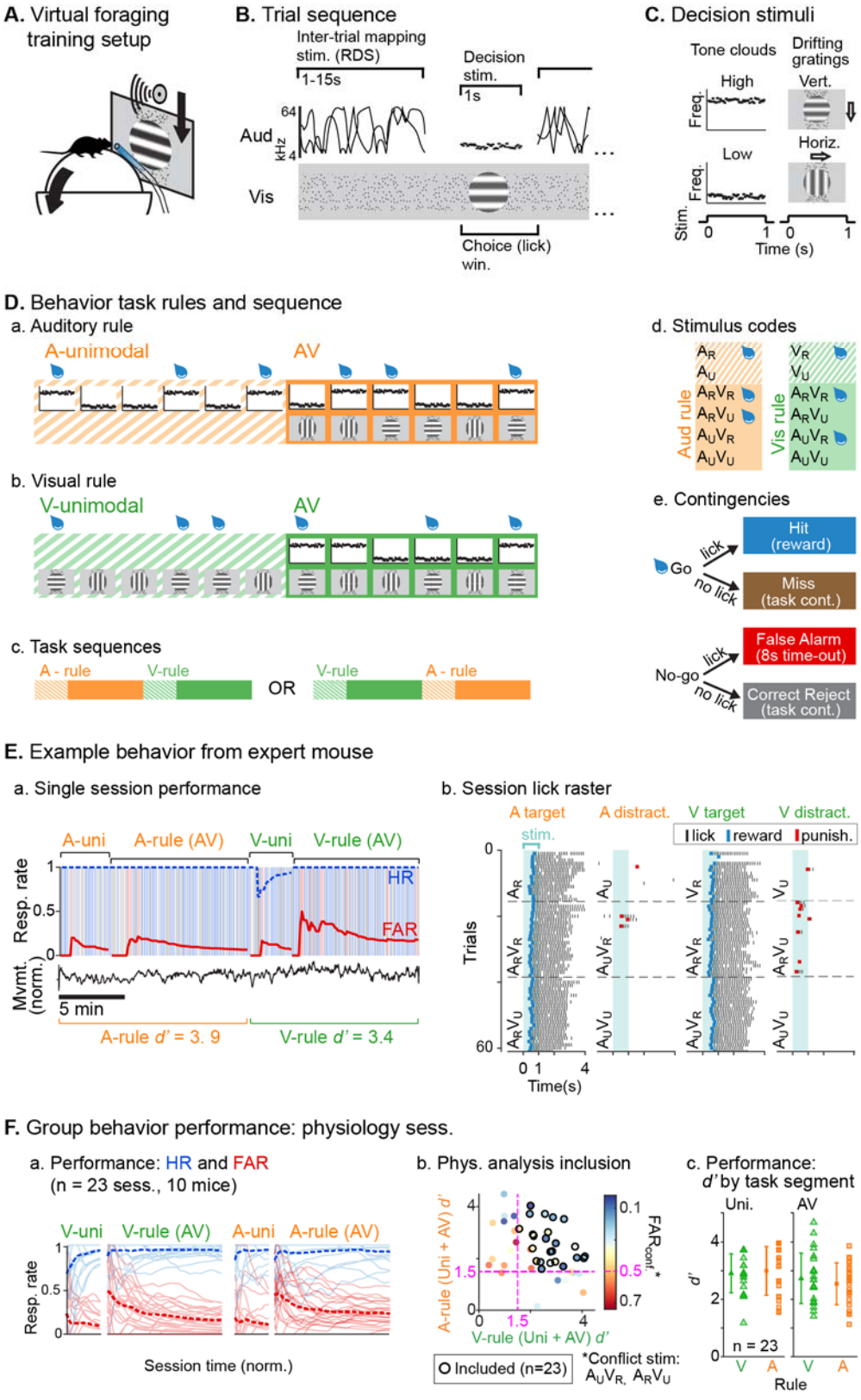


**Figure 1. Audiovisual rule-switching in mice.**

**A.** Virtual foraging environment: a head-fixed mouse runs on a floating spherical treadmill. Locomotion measured by treadmill movement controls auditory and visual stimulus presentation. A water spout in front of the mouse provides rewards. A lickometer records licking, which determines reward or punishment.
**B.** Trial sequence: during inter-trial intervals, a track of moving dots provides visual feedback for task progression while a random double sweep (RDS) auditory mapping stimulus is presented. Decision stimuli, either unimodal (auditory, A; visual, V) or bimodal (AV), are presented for 1 s. Choice window begins at decision stimulus onset, but trials with early licks (<0.3 s post-stimulus onset) are removed from subsequent analysis.
**C.** Decision stimuli comprised tone-clouds (TC; 8 or 17 kHz-centered) or drifting gratings (horizontal or vertical orientation). Each mouse is trained to lick for one auditory stimulus and one visual stimulus. Target/distractor stimulus identities were counterbalanced across mice for A- and V-rules.
**D.** Task sequences, attention cueing and reward contingencies. **a-b.** Behavioral sessions begin with a unimodal block, which cue the rule for the subsequent AV block. Water drops represent target stimuli, when mice have an opportunity for reward. **c.** Each session reflected one of two possible task sequences. **d.** Stimulus codes, for reference. **e.** Contingencies for water reward, time-out punishment, or task continuation.
**E.** Example behavior session. **a.** Hit rate (HR) and false alarm rate (FAR) across task blocks; trials and outcomes indicated by colored background bars. Mouse locomotion is shown below. **b.** Stimulus-onset aligned lick rasters for example session, organized by rule and target/distractor. Note that errors are typically false alarms on trials with 'conflict' stimuli: $A_UV_R$ in A-rule or $A_RV_U$ in V-rule.
**F.** Performance for all sessions included in subsequent physiology analysis. **a.** HR and FAR for all sessions organized by rule block; dashed lines indicate means. **b.** Performance metrics, showing dual inclusion filters: 1. sensitivity index *d'* performance index >1.5 for both A-rule and V-rule and 2. $FAR_{conf.}$<0.5 on conflict stimuli, as a critical test of modality-selective attention. **c.** *d'* is similar across task rules in unimodal and AV segments.

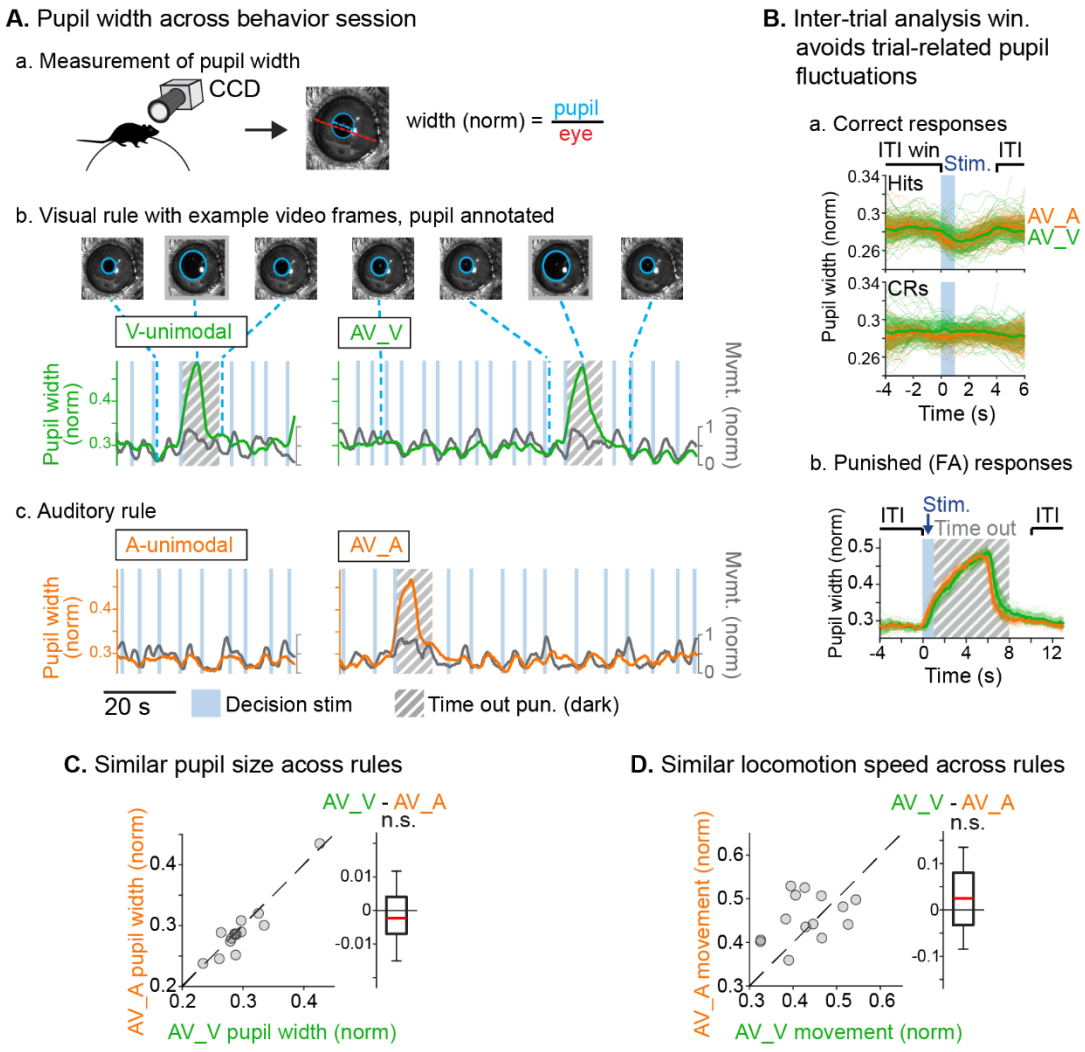


**Figure 2. Similar levels of arousal and movement during auditory and visual attention.**
**A.** Pupil size measurement. **a.** Left eye pupil recorded via CCD camera during the task. Pupil circumference (light blue) is tracked using automated video analysis; size is measured as pupil diameter over visible eye diameter. **b.** Example pupil video recorded during visual rule. Upper: annotated sample frames from times indicated by blue dashed lines. Lower: pupil width (green) and locomotion (gray) traces, with target stimuli and time out punishments indicated. Large fluctuations of pupil size occur during timeouts due to drop in light level (hashed gray background). **c.** Auditory rule from the same session.
**B.** Pupil size is measured during an inter-trial interval (ITI) window selected to capture engagement and arousal levels during each block and minimize influence from trial-related events such as rewards and timeouts. **a.** Pupil size decreases during hit trials due to reward administration. Correct reject trials (CRs; bottom) show no such decrease in running speed. **b.** Pupil size increases during timeout punishment when the recording chamber goes dark; ITI pupil size analysis window removes punishment-related fluctuations from analysis.
**C.** Pupil size is similar across V-rule bimodal and A-rule bimodal segments (pupillometry recorded in n = 14 sessions, 5 mice), suggesting similar levels of arousal and task engagement across rules.
**D.** Min-max-normalized locomotion is also similar across rules.

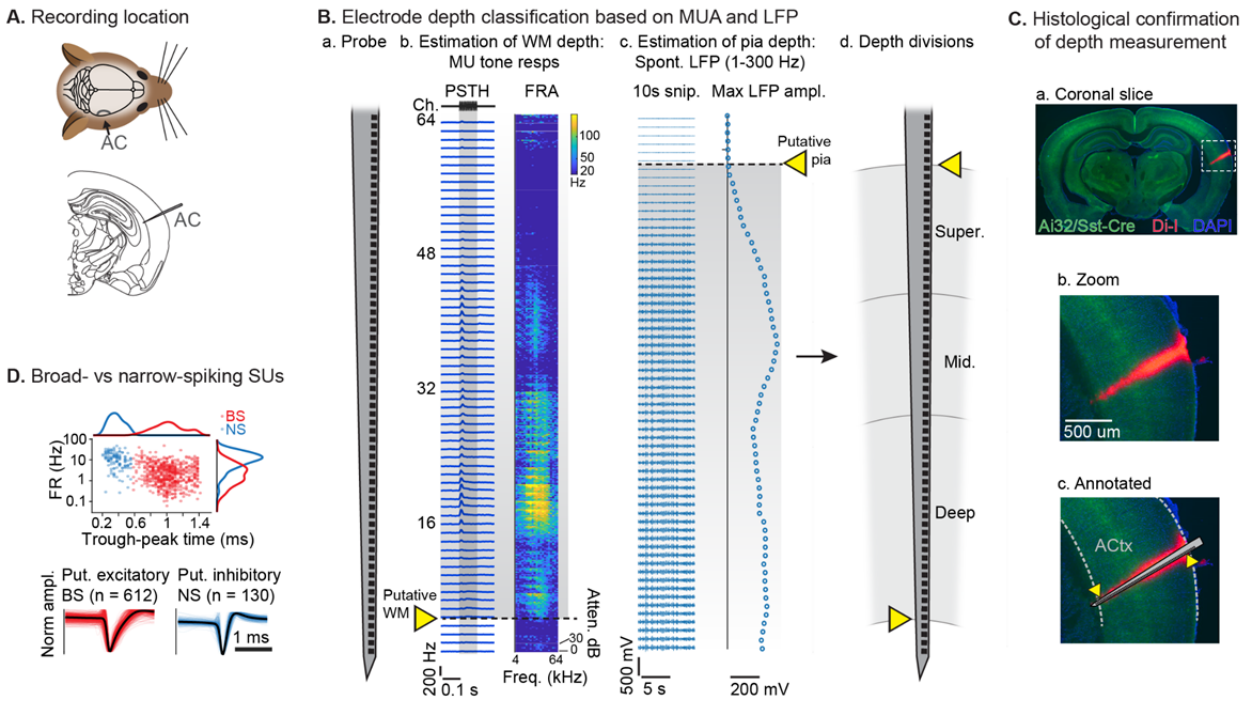


**Figure 3. Single unit recording and depth estimation in auditory cortex.**

**A.** Translaminar probes were used to record activity in right AC.

**B.** Physiological estimation of cortical depth. **a.** Linear 64-channel probe captures all activity in layers of AC. **b.** Example tone-evoked multi-unit (MU) sound responses by channel, providing a marker for the border of deep cortex and white matter (WM). Left: PSTH plots showing mean tone response by time. Right: frequency response area (FRA) shows mean response during tone stimulus by frequency/attenuation. MU responses poorly estimate the upper cortical boundary due to low somatic spiking activity in the superficial cortex. **c.** Local field potential (LFP; 1-300 Hz filtered) provides a marker for the upper cortex-pia boundary. Left: 10 s snippet of LFP by channel. Right: maximum LFP amplitude by channel,with putative pia location defined as the first deviation from probe-wise minimum LFP amplitude. **d.** Channels are assigned cortical depths based on fractional division of cortex into 'superficial', 'middle', and 'deep', with fractions based on supragranular, granular and infragranular anatomical divisions.

**C.** Histological confirmation of cortical depth estimation technique. **a.** Coronal slice showing DI-I probe track (red) in right AC. Green: eYFP fluorescence from Ai32/Sst-Cre mouse strain. Blue: DAPI stain to visualize cell bodies. **b.** Zoomed area indicated by dashed rectangle in a. **c.** Probe overlay and WM/pia boundaries. Yellow arrows indicate locations of physiologically-determined cortical span from B., showing close correspondence with Di-I probe track.

**D.** Sorted SU waveforms were divided into narrow-spiking (putative fast-spiking inhibitory) and broad-spiking (putative excitatory) based on a waveform trough-peak time boundary of 0.6 ms.

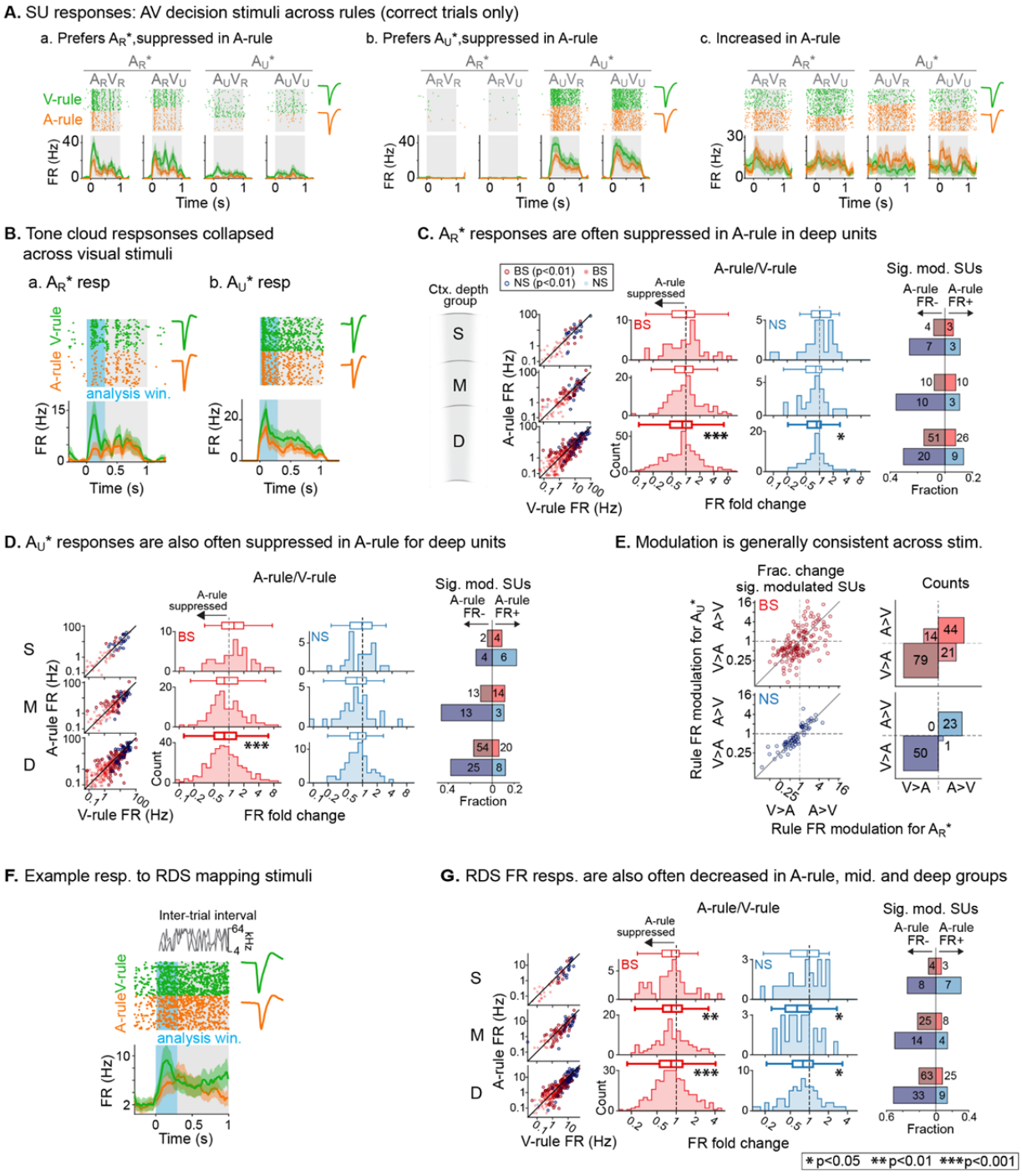


**Figure 4. Net suppression of sound-evoked firing rates during auditory attention.**

**A.** Example SU responses to physically identical AV stimuli across task rules ($A_R$* = $A_RV_R$ and $A_RV_U$ collapsed; $A_U$* = $A_UV_U$ and $A_UV_R$ collapsed;). **a.** Preference for $A_R$* tone-cloud, suppressed in A-rule relative to V-rule. **b.** Preference for $A_U$*tone cloud, suppressed in A-rule. **c.** Moderately enhanced FR in A-rule.

**B.** Example SU responses to **a.** $A_R$* and **b.** $A_U$* tone clouds (TCs). Early sensory-driven response analysis window (0 - 0.3 s) shown in light blue.

**C.** Group data: responses to TCs rewarded in A-rule ($A_R$*) between rules by unit type and depth. Scatter plots (left) show FR across rules. Red: BS units. Blue: NS units. Outlined: significantly modulated units, paired t-test, Benjamini-Hochberg false discovery rate (FDR)-adjusted, q = 0.01. Fold change histograms show A-rule FR divided by V-rule FR for all units; bins to the left of 1 (dashed line) indicate FR suppression in A-rule. Box plots above histograms: central line: median; box edges: 25th and 75th percentiles; whiskers: data points not considered outliers. Asterisks indicate FDR-adjusted (q=0.05, n=6 tests) p-values from paired Wilcoxon signed-rank tests of mean FRs across rules; no asterisk: not significant (p>0.05).
Right: fractions of significantly modulated units (inclusion as described above) over total. Darker colors indicate fractions with significantly suppressed FRs in A-rule; lighter colors indicate enhanced FRs in A-rule.
**D.** Responses to TCs unrewarded in A-rule ($A_U$*). All conventions as in C.
**E.** Comparison of unit FR modulation by rule between $A_R$* (abscissa) and $A_U$* (ordinate). Top: BS units, bottom: NS units. Scatter plots (left) show all units with significant rule modulation for $A_R$*, $A_U$* or both. Modulation values <1 indicate suppressed FR response in A-rule. Note the increased density of units below 1 for BS and NS units. Right: counts of units by direction of FR rule modulation. Most units lie in quadrants with similar direction of modulation across stimuli, suggesting that attentional effects on FR are not frequency or stimulus identity-dependent.
**F**. Example SU response to the onset of the random double sweep (RDS) mapping stimulus, showing analysis window for calculating FR (0-0.3 s, blue).
**G.** Group data for RDS FR modulation across rules by depth and BS/NS classifications. All conventions as in C.

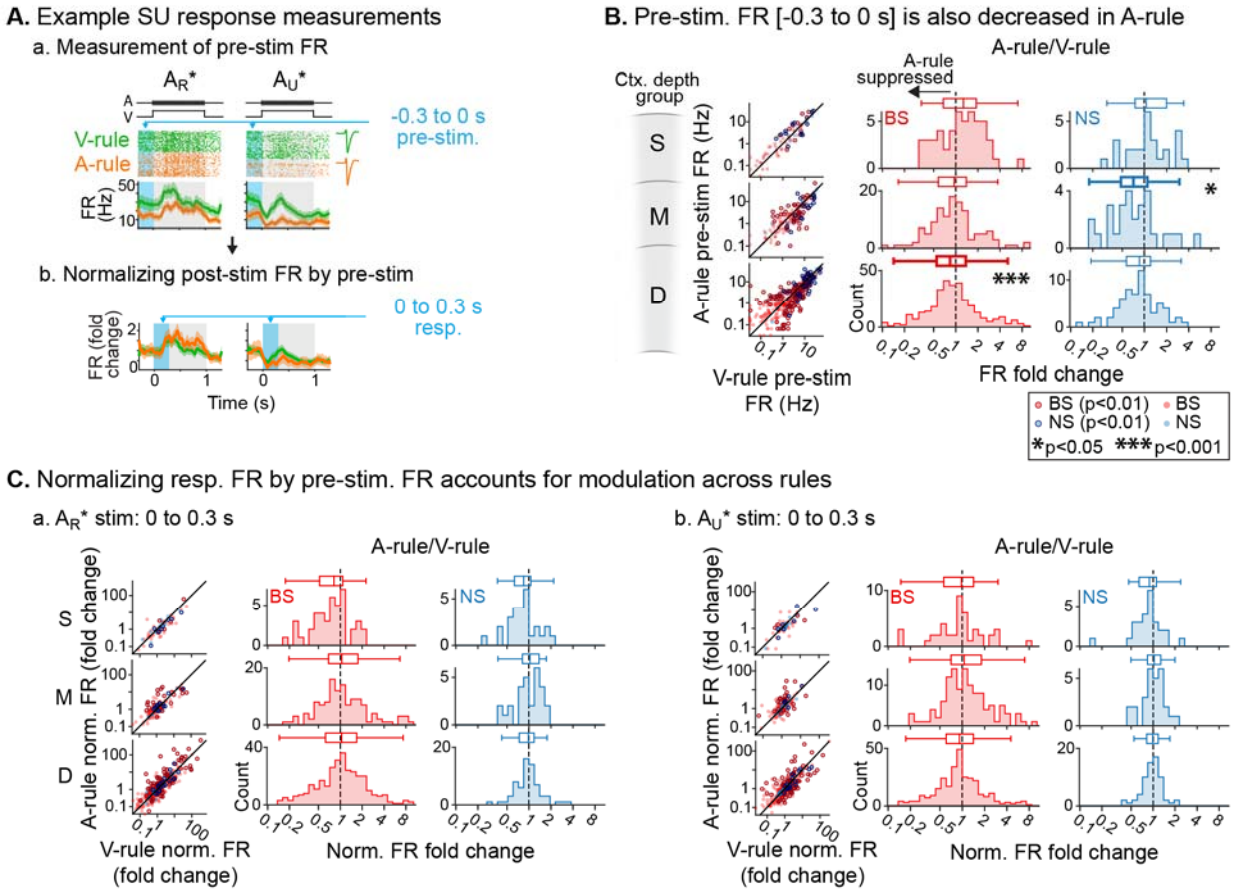


**Figure 5. Attention-related modulation of sound-evoked responses largely reflects pre-stimulus activity changes.**

**A.** Example pre-stimulus FR measurement, and normalization of post-stimulus response. **a.** Raw FR by condition and stimulus. Pre-stimulus analysis window shown in blue (-0.3 to 0 s). **b.** Normalized FRs (FR divided by mean pre-stimulus FR).

**B.** Group data: pre-stimulus onset FR compared across rules, with data organized by depth (S = superficial, M = middle, D = deep) and BS/NS (red/blue). Conventions as in Figure 4. Scatter plots (left) show individual units, with significantly modulated units outlined (paired t-test, Benjamini-Hochberg FDR-adjusted, q = 0.01). Difference histograms show A-rule/V-rule for all units shown in scatters; fold change <1 indicates suppression during the A-rule. As in Figure 4, asterisks represent *p*-values from FDR-adjusted paired Wilcoxon signed-rank tests on each group (q = 0.05, n = 6 tests). Absence of asterisk: not significant.

**C.** Group data: response as fold change, normalized by pre-stimulus FR. Conventions as in B. and Figure 4. After accounting for pre-stimulus modulation, effects of rule on FR are abolished.

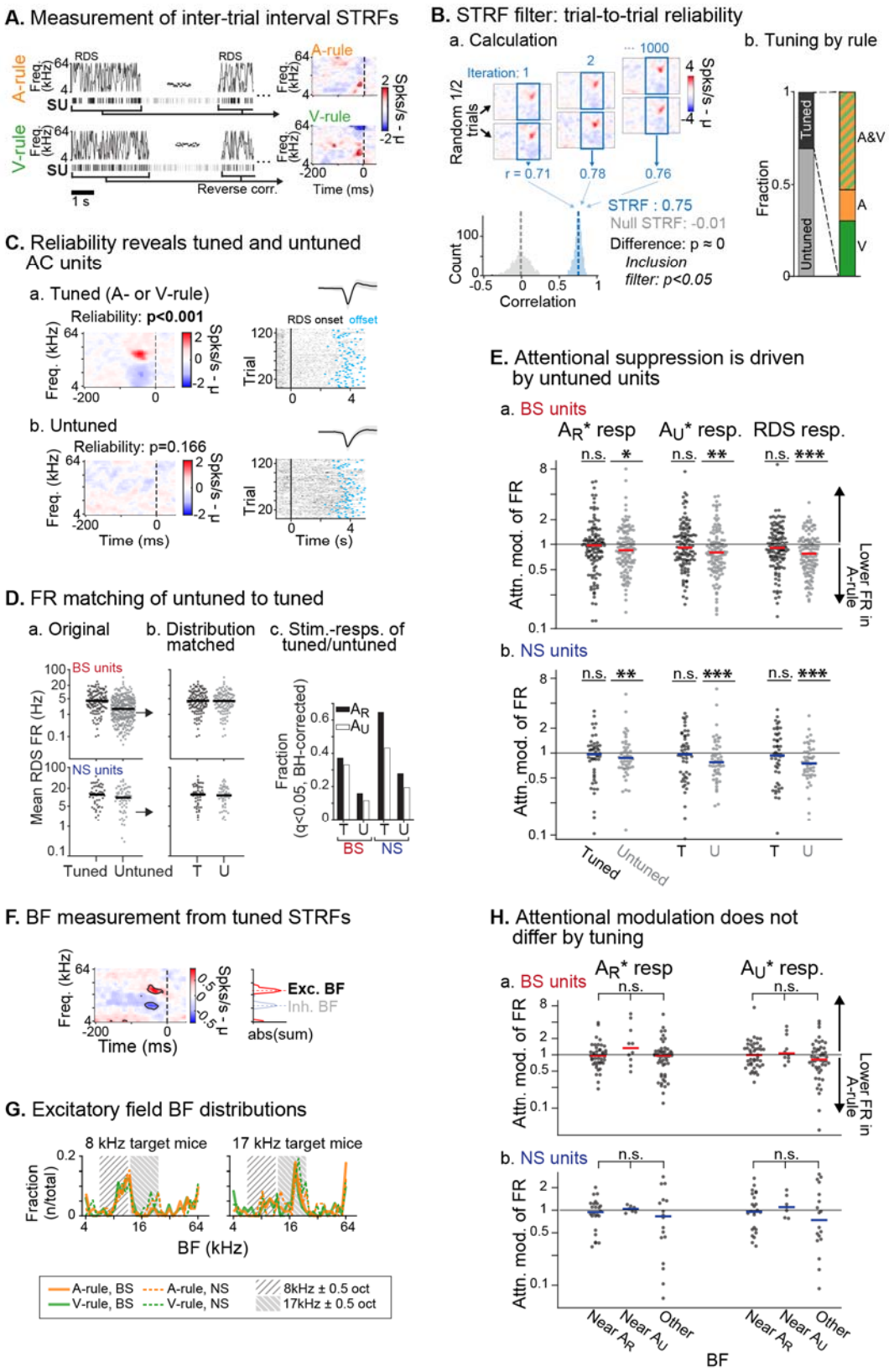


**Figure 6. Attentional modulation of spike rate is driven by neurons without STRF tuning.**

**A.** Spectrotemporal receptive fields (STRFs) for A-rule and V-rule were calculated from spikes during the inter-trial-interval random-double sweep (RDS) stimulus using standard reverse correlation methods.

**B.** STRF reliability as a measure for tuning. **a.** Reliability was measured through correlations of randomly subsampled halves of all RDS presentations, repeated 1000 times. A *p*-value was calculated empirically through comparison of correlation value distributions from the actual STRF and a null STRF, generated from random circular shuffling of spike trains relative to stimulus. **b.** Left: fraction of tuned and untuned units. Right: fractions of units with STRF tuning in both A- and V-rules (AV), A-rule only (A) or V-rule only (V).
**C.** Trial-to-trial reliability metric separates AC units into those with tuned STRFs (**a.**) and untuned STRFs (**b.**).
**D.** To control for activity level-driven effects, the larger group with untuned STRFs (n = 345, 64 for BS, NS) is matched for sample size and firing rate to the group with tuned STRFs (n = 121, 51 for BS, NS). **a.** Mean FR during RDS mapping stimulus by tuned and untuned groups. **b.** FR distribution-matched groups. **c.** Tuned group contains a larger share of decision stimulus-responsive units compared with untuned (for both $A_R$* and $A_U$* TCs). Stimulus-responsiveness is defined as a significant FR difference between 0.3 - 0 s pre-stimulus window and 0 - 0.3 s post-stimulus window, paired t-test, Benjamini-Hochberg FDR-adjusted, q = 0.01.
**E.** Untuned unit group is suppressed during auditory attention, while tuned unit group is not. **a.** Attentional modulation of broad-spiking unit responses for task decision stimuli (left: $A_R$*; right: $A_U$*) and RDS mapping stimuli (right). Paired Wilcoxon signed-rank between mean FR in A-rule and V-rule, FDR-corrected at q = 0.05 (n = 3 comparisons per group). Asterisks indicated FDR-adjusted *p*-values. **b.** Narrow-spiking, conventions as in a. Asterisks indicate significance: * p<0.05; **p<0.01; *** p<0.001.
**F**. Measurement of best frequency (BF) from tuned STRF group, based on peaks of absolute values of significant time-frequency bins summed across time (-100 ms to 0 window). Significant time-frequency bins (p<0.01) determined by comparison of observed STRF values with distribution of values from spike time-shuffled null STRF.
**G.** BFs of excitatory STRF fields show that AC units are preferentially tuned near the center frequency of the target (rewarded) TC.
**H.** Attentional modulation by BF of tuned units: tuned near $A_R$ (BF ± 0.5 octaves from TC center), $A_U$, or tuned to frequency outside either band. Response modulation does not differ by BF tuning for any comparison ($A_R$ or $A_U$ response and BS or NS units; Kruskal-Wallis Test; BS: all p> 0.11, NS: all p>0.81, FDR-adjusted).

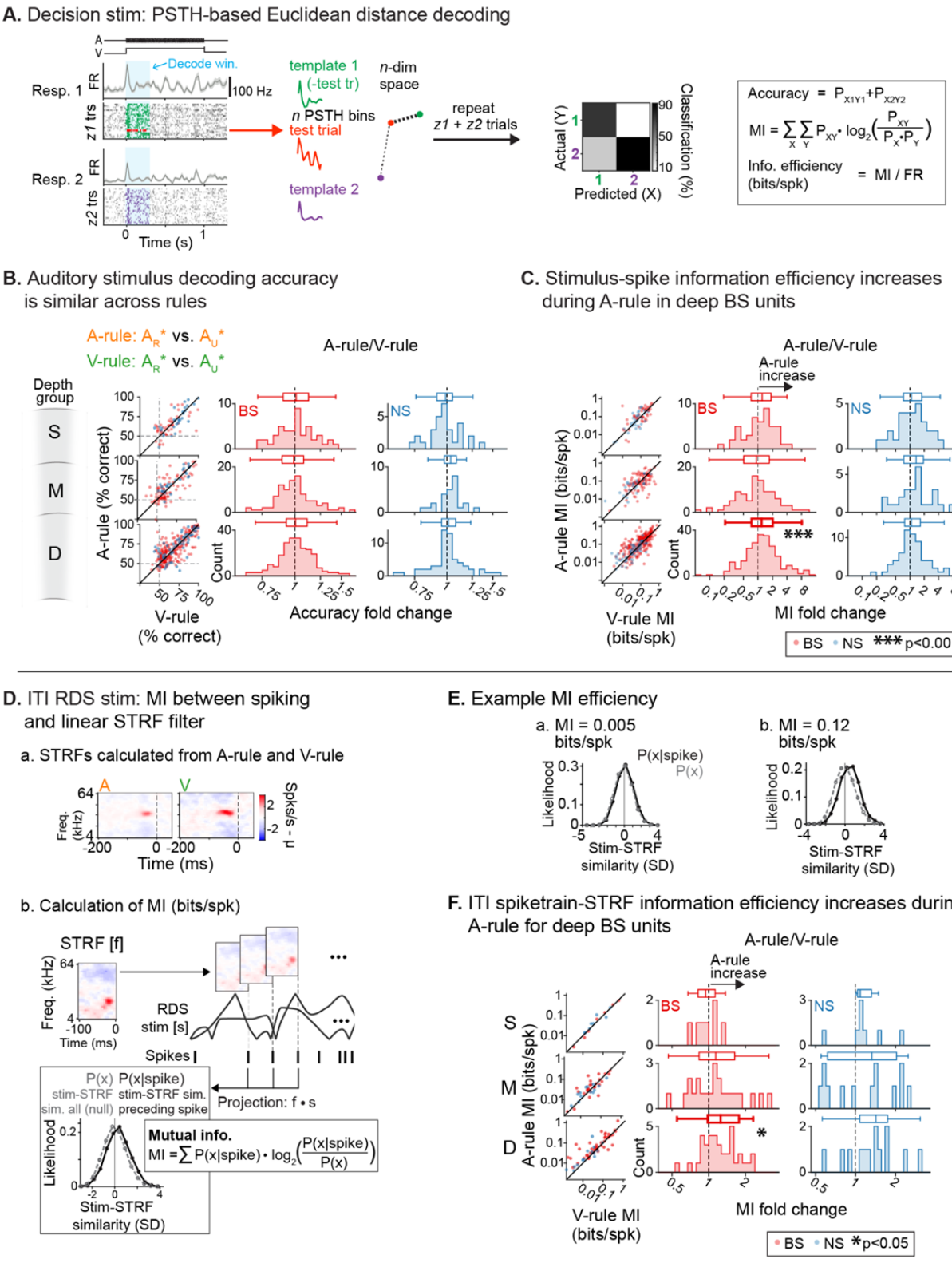


**Figure 7. Auditory attention increases sound encoding efficiency in deep layer broad-spiking units.**

**A.** PSTH-based spiketrain decoding analysis. Time-binned responses for each single trial (test trial; red) are compared to PSTHs (templates; green, purple) reflecting responses to each stimulus averaged across all other trials. Trials are classified as belonging to the template nearest to the test trial in *n*-dimensional Euclidean space

(*n* = number of PSTH bins). A confusion matrix (right) reflecting predicted/actual outcomes for all trials is used to calculate accuracy, mutual information (MI; bits) and encoding efficiency (bits/spk).

**B.** Decoding accuracy of auditory stimulus identity, compared across attentional states. Decoder setup mimics task faced by mice in the A-rule: discrimination between $A_R$* and $A_U$* tone cloud identities. Results represent average of two decoder runs, which differ in their paired visual stimulus, but yield highly similar results: $A_RV_R$ vs. $A_UV_R$ and $A_RV_U$ vs. $A_UV_U$ (see **Figure 7 - figure supplement 2** for separate presentation of these data). Conventions as in Figure 4C. and elsewhere. Scatter plots represent decoder accuracy from individual units; dashed lines show chance level (50%).

Histograms show raw unit counts for A-rule/V-rule fold change. S = superficial; M = middle; D = deep. No change in accuracy is observed across rules.

**C.** Stimulus-spike information efficiency (bits/spike, calculation shown in A.) for PSTH-based decoding increases for deep broad-spiking units during auditory attention. Conventions as in B.

**D.** Measuring encoding changes across attentional states for inter-trial interval (ITI) mapping stimuli. **a.** Example STRFs calculated from ITIs of A-rule and V-rule from the same SU. **b.** Estimation of mutual information efficiency of ITI RDS stimuli: the STRF is convolved with the windows of the RDS stimulus to define two distributions of relative STRF-stimulus similarity values: 1. P(x|spike), from time windows preceding a spike, and 2. P(x), a null distribution from non-overlapping time windows tiling the full stimulus duration. Information encoding efficiency is calculated as shown, reflecting the divergence between these distributions, which increases when spiking preferentially occurs during periods of higher stimulus-STRF similarity. MI values are calculated from STRFs in A-rule and V-rule separately.

**E.** Example of spiketrain-STRF encoding efficiency from two SUs: low (**a**) and high (**b**) bits/spk examples.

**F.** Comparison of spiketrain-STRF encoding efficiency across rules, showing increased encoding efficiency in A-rule for deep broad-spiking units.

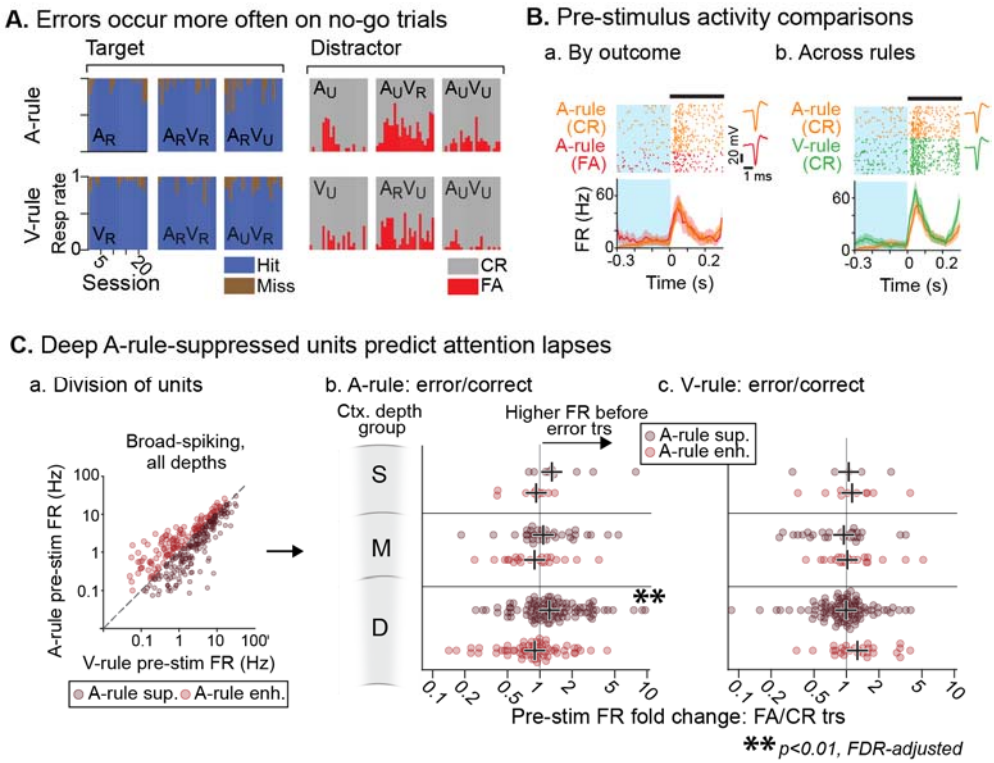


**Figure 8. Attentionally-suppressed units predict behavior performance during auditory attention.**

**A.** Summary of behavioral outcomes by session (n = 23, 10 mice), organized by task stimulus. Bar sequence follows chronology of experiments. Error trials are predominantly false alarms (FAs). To allow sufficient trials for measurement of activity levels across behavioral outcomes, subsequent analysis focuses on analysis of false alarms (FAs) versus correct rejects (CRs).

**B.** Example unit showing behavioral outcome- and rule-dependent FR modulation. Pre-stimulus FR analysis window (-0.3 to 0 s) shown in blue. **a.** Pre-stim. activity for A-rule FA trials (red) is elevated relative to CR trials (orange). **b.** In the same unit, pre-stim. activity is elevated in V-rule CR trials (green) relative to A-rule CR trials (orange).

**C.** Division of units into A-rule-suppressed and A-rule-enhanced groups reveals suppression of activity as a neural signature of correct task performance. **a.** BS units from sessions with ≥ 10 FA and CR trials are divided into A-rule suppressed and A-rule enhanced groups. **b.** Deep units that are suppressed during auditory attention relative to visual show higher firing rates on A-rule error trials relative to correct trials ($p = 0.0098$, paired Wilcoxon signed-rank test, FDR-adjusted for n = 6 tests). Median of group indicated by black cross. No such trend exists for the A-rule-enhanced population. **c.** Pre-stimulus activity does not predict V-rule behavioral outcomes in the same groups, suggesting that AC activity suppression is related to performance on sound but not visual stimulus discrimination (all $p>0.68$, paired Wilcoxon signed-rank test, FDR-adjusted for n = 6 tests).

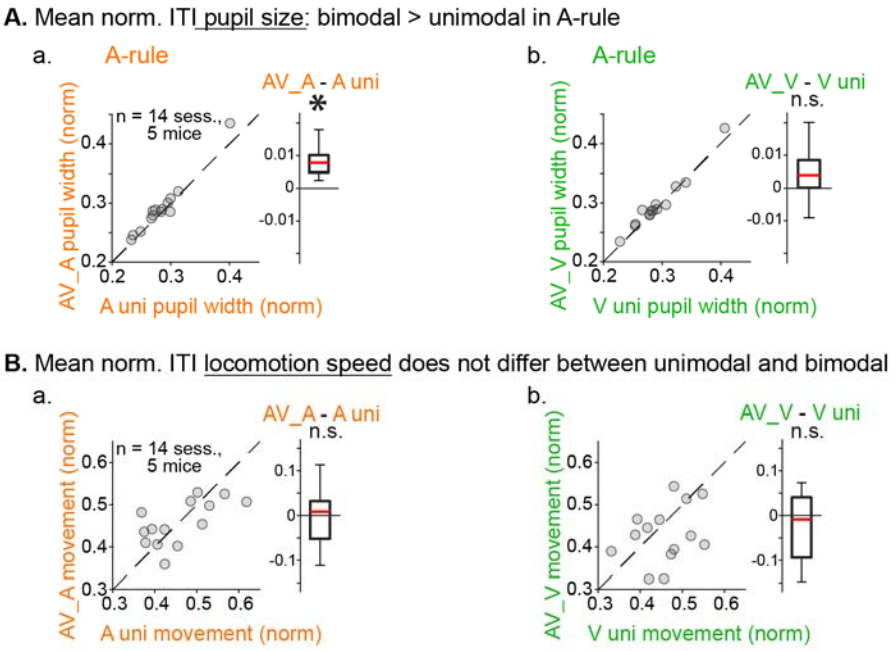


**Figure 2 - figure supplement 1. Pupil size and locomotion compared between unimodal and bimodal blocks.**

**A.** Pupil size, measured as described in Figure 2, compared within rule between unimodal and bimodal task segments. **a.** A-rule: unimodal A-only pupil size is smaller than bimodal AV pupil (A-rule unimodal: 0.28 ±0.04 mean norm. pupil width ± SD, A-rule bimodal: 0.29 ±0.05; Z = -2.6, p = 0.009), possibly due to the presence of drifting grating visual stimulation or increased task difficulty in the bimodal segment. **b.** V-rule: unimodal V-only pupil size trends toward smaller than that in bimodal AV pupil, but does not reach significance after multiple comparisons correction (V-rule unimodal: 0.29 ±0.04, V-rule bimodal: 0.30 ±0.05; Z = -2.0, p = 0.062).

**B.** Min-max-normalized locomotion is similar across unimodal and bimodal task segments. **a.** A-rule: locomotion during unimodal is comparable to locomotion during bimodal (A uni: 0.46 +/- 0.08, A bimodal: 0.46 +/- 0.05; Z = 0.282, p = 0.78). **b.** V-rule: locomotion during unimodal is comparable to locomotion during bimodal (V uni: 0.46 +/- 0.06, V bimodal: 0.43 +/- 0.07; Z = 1.224, p = 0.33). All *p*-values FDR adjusted for n = 3 comparisons (including A-rule bimodal vs. V-rule bimodal in Figure 2). Difference box plots: central line: median; box edges: 25th and 75th percentiles; whiskers: data points not considered outliers.

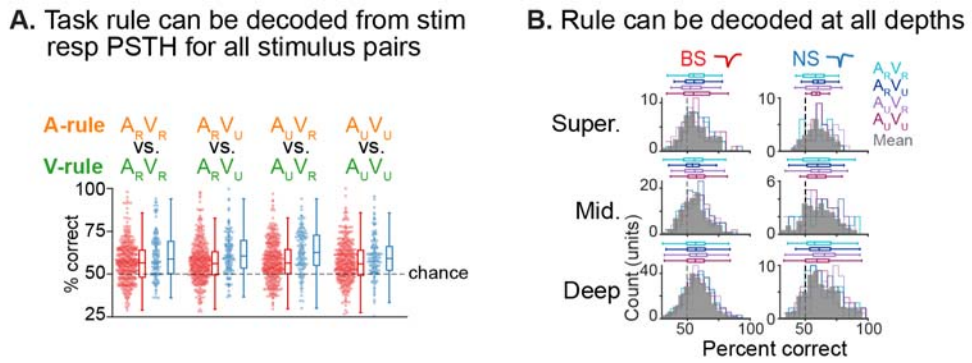


**Figure 7 - figure supplement 1. Decoding of task rule from stimulus response PSTHs.**

**A.** Decoding of rule identity from PSTH responses using Euclidean distance-based classifier. Responses to each of the four AV stimuli in the A-rule vs responses to the same stimuli in the V-rule were decoded based on 0-0.3 s stimulus onset window as described in **Figure 7** and in text; PSTHs were constructed with 30 ms bins. Each dot represents SU decoding accuracy (% correct); BS units in red, NS in blue. Chance decoding is indicated by the dashed line. A minimum 1 Hz FR response to both stimuli in the decode was required for inclusion. Box plots as before: central line indicates median, box edges indicate 25th and 75th percentiles, whiskers extend to data points not considered outliers. A repeated-measures ANOVA, showed no difference of stimulus or BS/NS group in decoding (stimulus [within-subject]: $F(2.8,1050.6) = 0.73$, $p = 0.52$; BS/NS [between-subject]: $F(2.8,1050.6) = 0.68$, $p = 0.55$; Greenhouse-Geisser corrected).

**B.** Decoding by depth group. Median decoding by stimulus indicated as colored horizontal box plots (as described above). Gray distributions represent decoding by depth, averaged across stimulus types. Left: broad-spiking units, right: narrow-spiking. A two-way ANOVA shows a significant effect of depth ($F(2,579) = 3.75$, $p = 0.024$), BS/NS $F(1,579) = 27.3$, $p \approx 0$), but not their interaction ($F(2,579) = 0.61$, $p = 0.55$). Post-hoc comparison by depth showed that decoding was significantly better in the deep compared to the superficial groups and the NS compared to BS unit groups. See **Figure 7 - source data 1** for additional statistics.

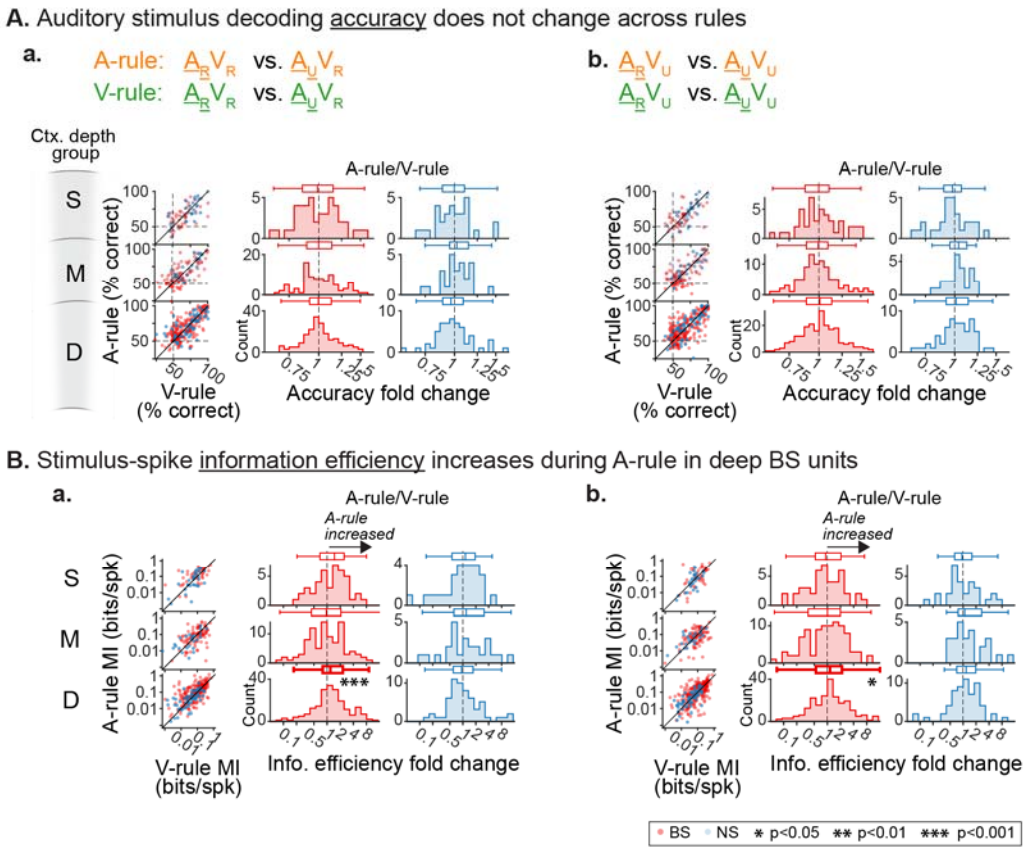


**Figure 7 - figure supplement 2. Decoding and information efficiency changes across rules are similar across visual stimulus pairings.**

Decoding of auditory stimulus identity, compared across attentional states, similar to **Figure 7A-C** but showing decoder runs separated by paired visual stimuli.

**A.** Accuracy of discrimination between $A_R$ and $A_U$ stimuli across rules. **a.** $A_RV_R$ vs. $A_UV_R$: no difference across rules (all p≥0.36, all |Z|≤0.92, paired WSR test, see **Table. S8**). **b.** $A_RV_U$ vs. $A_UV_U$: no difference across rules (all p≥0.24, all |Z|≤1.17) . Conventions as in Figure 4C. and elsewhere. Scatter plots represent decoder accuracy from individual units; dashed lines show chance level (50%). Histograms show raw unit counts for A-rule/V-rule fold change. S = superficial; M = middle; D = deep.

**B.** Stimulus-spike information efficiency (bits/spike, calculation shown in A.) for PSTH-based decoding increases for deep broad-spiking units during auditory attention. Conventions as in A. **a.** $A_RV_R$ vs. $A_UV_R$: deep group shows increased information efficiency in A-rule (Deep BS: p = 5.8e-4, Z = 1.2, paired WSR test, *p*-values FDR-corrected; see **Figure 7 - source data 4.** for other groups) **b.** $A_RV_U$ vs. $A_UV_U$: deep BS group also shows increase in A-rule (p = 0.036, Z = 1.15).

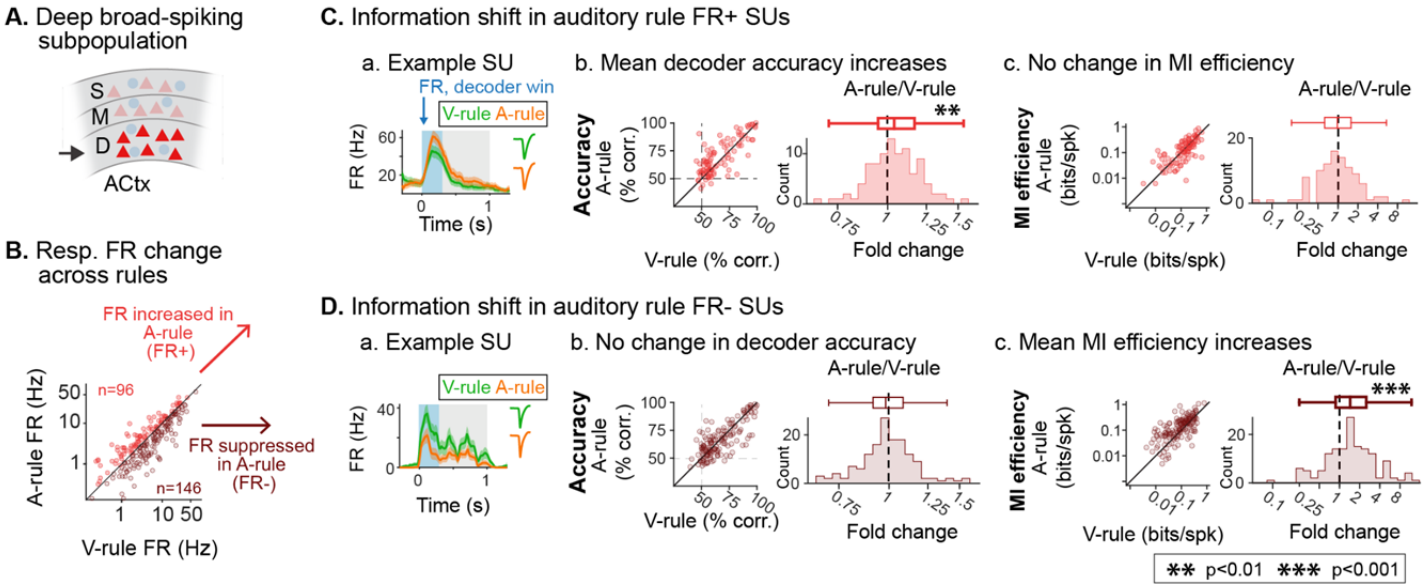


**Figure 7 - figure supplement 3. Deep AC A-rule information efficiency increases are driven by FR suppressed units.**

**A.** Analysis is restricted to unit group exhibiting information efficiency modulation: deep broad-spiking units (see **Figure 7**).

**B.** Decoder-based analyses are split into subpopulations showing increased stimulus responses in the A-rule (FR+; n = 96, 39.7%) or decreased responses in the A-rule (FR-; n = 146, 60.3%).

**C.** Restriction of analysis to FR+ units. **a.** Example unit, decision stimulus-evoked PSTH response during A-rule (orange) and V-rule (green). b**.** PSTH-based decoder accuracy for FR+ group shows a significant increase in A-rule. Left, scatter showing accuracy across rules, points represent individual SUs. Right, difference histograms. For succinctness, decoder accuracy shown is the mean of $A_RV_R$ vs. $A_UV_R$ and $A_RV_U$ vs. $A_UV_U$ decoding across rules. **c.** No change is observed for MI efficiency across rules for FR+ deep BS units (conventions as in **b**).

**D.** Similar analyses for FR- units. **a.** Example unit with reduced response in A-rule. **b.** Decoder accuracy does not change across rules despite loss of spikes in the FR- subpopulation. **c.** MI efficiency increases are driven by the FR- population, directly linking changes in information rate to changes in activity levels across attentional state. See text for statistics.

**Figure 4 - source data 1.A**

**Analysis: AR* response FR across rules**

| *Group* | *n* | *A-rule mean FR (Hz)* | *±SD* | *V-rule mean FR (Hz)* | *±SD* | *WSR Z-value* | *p-value* | *FDR-adjusted p-value* | *Median fold change across rules (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 49 | 3.78 | 6.54 | 3.58 | 5.26 | -0.49 | 0.62 | 0.744 | -- |
| Super. NS | 27 | 13.49 | 15.52 | 14.02 | 18.1 | 0.22 | 0.83 | 0.83 | -- |
| Mid. BS | 127 | 6.48 | 14.49 | 5.88 | 9.19 | 1.03 | 0.3 | 0.45 | -- |
| Mid. NS | 28 | 13.11 | 10.6 | 16.14 | 11.17 | 2.23 | 0.026 | 0.052 | -- |
| Deep BS | 333 | 5.26 | 7.83 | 6.06 | 8.84 | 4.08 | 4.60E-05 | 2.76E-04 | 0.89 |
| Deep NS | 66 | 19.87 | 17.84 | 23.36 | 20.68 | 2.9 | 0.0038 | 0.0114 | 0.87 |

**Figure 4 - source data 1.B**

**Analysis: AU* response FR across rules**

| *Group* | *n* | *A-rule mean FR (Hz)* | *±SD* | *V-rule mean FR (Hz)* | *±SD* | *WSR Z-value* | *p-value* | *FDR-adjusted p-value* | *Median fold change across rules (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 46 | 4.39 | 7.8 | 3.44 | 5.69 | -1.24 | 0.21 | 0.252 | -- |
| Super. NS | 27 | 13.72 | 11.74 | 13.9 | 10.47 | 0.14 | 0.89 | 0.89 | -- |
| Mid. BS | 122 | 5.78 | 11.48 | 5.44 | 7.07 | 1.38 | 0.17 | 0.252 | -- |
| Mid. NS | 28 | 12.73 | 9.31 | 16.4 | 11.69 | 2.3 | 0.021 | 0.063 | -- |
| Deep BS | 321 | 4.69 | 7.58 | 5.82 | 10.01 | 5.11 | 3.30E-07 | 1.98E-06 | 0.81 |
| Deep NS | 66 | 17.39 | 14.11 | 19.24 | 13.78 | 1.91 | 0.056 | 0.112 | -- |

**Figure 4 - source data 2.**

**Analysis: RDS response FR across rules**

| *Group* | *n* | *A-rule mean FR (Hz)* | *±SD* | *V-rule mean FR (Hz)* | *±SD* | *WSR Z-value* | *p-value* | *FDR-adjusted p-value* | *Median fold change across rules (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 45 | 3.05 | 5.54 | 2.69 | 3.32 | 1.44 | 0.15 | 0.18 | -- |
| Super. NS | 23 | 9.21 | 8.23 | 9.8 | 5.97 | 0.52 | 0.61 | 0.61 | -- |
| Mid. BS | 112 | 3.66 | 5.95 | 4.17 | 5.33 | 3.09 | 0.002 | 0.006 | 0.85 |
| Mid. NS | 28 | 8.55 | 7.68 | 11.89 | 8.87 | 2.69 | 0.0072 | 0.0144 | 0.65 |
| Deep BS | 309 | 3.37 | 4.25 | 4.11 | 5.32 | 5.2 | 2.00E-07 | 1.20E-06 | 0.84 |
| Deep NS | 64 | 13.55 | 11.44 | 15.77 | 12.54 | 2.45 | 0.014 | 0.021 | 0.8 |

**Figure 5 - source data 1.**

**Analysis: Pre-stimulus FR across rules**

| *Group* | *n* | *A-rule mean FR (fold change)* | *±SD* | *V-rule mean FR (fold change)* | *±SD* | *WSR Z-value* | *p-value* | *FDR-adjusted p-value* | *Median fold change across rules (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 51 | 3.15 | 5.57 | 2.36 | 3.25 | -1.61 | 0.11 | 0.132 | -- |
| Super. NS | 27 | 9.59 | 8.68 | 8.99 | 7.68 | -1.06 | 0.29 | 0.29 | -- |
| Mid. BS | 127 | 3.99 | 7.23 | 4.1 | 5 | 2.01 | 0.045 | 0.072 | -- |
| Mid. NS | 28 | 10.46 | 9.63 | 13.44 | 10.95 | 2.48 | 0.013 | 0.039 | 0.71 |
| Deep BS | 336 | 3.44 | 4.55 | 4.07 | 5.37 | 4.49 | 7.00E-06 | 4.20E-05 | 0.87 |
| Deep NS | 66 | 13.67 | 11.75 | 15.02 | 12.55 | 1.98 | 0.048 | 0.072 | -- |

**Figure 5 - source data 2.A**

**Analysis: AR* baseline-adjusted FR across rules**

| *Group* | *n* | *A-rule mean FR (fold change)* | *±SD* | *V-rule mean FR (fold change)* | *±SD* | *WSR Z-value* | *p-value* | *FDR-adjusted p-value* | *Median fold change across rules (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 44 | 3.7 | 9.92 | 4.64 | 9.64 | 2.37 | 0.018 | 0.075 | -- |
| Super. NS | 27 | 2.34 | 3.08 | 3.26 | 5.59 | 2.23 | 0.025 | 0.075 | -- |
| Mid. BS | 125 | 2.69 | 4.47 | 3.23 | 7.55 | 0.06 | 0.95 | 0.95 | -- |
| Mid. NS | 28 | 2.52 | 4.02 | 3.09 | 6.89 | -0.55 | 0.58 | 0.708 | -- |
| Deep BS | 310 | 8.51 | 45.06 | 8.95 | 53.43 | -1.27 | 0.21 | 0.42 | -- |
| Deep NS | 66 | 2.54 | 4.31 | 2.35 | 2.91 | 0.54 | 0.59 | 0.708 | -- |

**Figure 5 - source data 2.B**

**Analysis: AU* baseline-adjusted FR across rules**

| *Group* | *n* | *A-rule mean FR (fold change)* | *±SD* | *V-rule mean FR (fold change)* | *±SD* | *WSR Z-value* | *p-value* | *FDR-adjusted p-value* | *Median fold change across rules (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 43 | 1.92 | 1.92 | 2.3 | 2.95 | 1.06 | 0.29 | 0.435 | -- |
| Super. NS | 27 | 2.57 | 3.49 | 4.59 | 12.67 | 2.07 | 0.039 | 0.144 | -- |
| Mid. BS | 118 | 2.23 | 4.04 | 1.62 | 2.23 | -1.52 | 0.13 | 0.26 | -- |
| Mid. NS | 28 | 2 | 2.82 | 1.87 | 2.25 | -0.48 | 0.63 | 0.63 | -- |
| Deep BS | 306 | 4.46 | 23.66 | 4.24 | 21.24 | 1.98 | 0.048 | 0.144 | -- |
| Deep NS | 66 | 2.05 | 2.72 | 2.46 | 4.99 | 0.61 | 0.54 | 0.63 | -- |

1148

**Figure 6 - source data 1.A**

**Analysis: Stimulus response modulation across rules by tuned/untuned, BS cells**

| *Group* | *Stimulus* | *n* | *Mean FR resp across rules (fold change)* | *±SD* | *WSR Z-value (vs 1 [no modulation by rule])* | *p-value* | *FDR-corrected p-value (if any p<0.05)* |
|---|---|---|---|---|---|---|---|
| Tuned | AR* | 121 | 0.97 | 0.99 | -1.30 | 0.193 | 0.193 |
| | AU* | 120 | 0.90 | 0.98 | -1.57 | 0.117 | 0.176 |
| | RDS | 121 | 0.90 | 0.90 | -1.72 | 0.086 | 0.176 |
| Untuned | AR* | 120 | 0.89 | 0.97 | -2.27 | 0.023 | **0.023** |
| | AU* | 118 | 0.76 | 1.51 | -2.85 | 4.36E-03 | **6.54E-03** |
| | RDS | 121 | 0.77 | 0.51 | -4.46 | 8.37E-06 | **2.51E-05** |

***Note: fold change is A-rule / V-rule***

**Figure 6 - source data 1.B**

**Analysis: Stimulus response modulation across rules by tuned/untuned, NS cells**

| *Group* | *Stimulus* | *n* | *Mean FR resp across rules (fold change)* | *±SD* | *WSR Z-value (vs 1 [no modulation by rule])* | *p-value* | *FDR-corrected p-value* |
|---|---|---|---|---|---|---|---|
| Tuned | AR* | 51 | 0.97 | 0.64 | -1.23 | 0.219 | 0.220 |
| | AU* | 51 | 0.96 | 0.71 | -0.83 | 0.404 | 0.404 |
| | RDS | 51 | 0.93 | 0.74 | -1.04 | 0.298 | 0.298 |
| Untuned | AR* | 51 | 0.84 | 0.78 | -2.62 | 8.92E-03 | **8.92E-03** |
| | AU* | 51 | 0.71 | 0.91 | -3.29 | 1.00E-03 | **1.50E-03** |
| | RDS | 51 | 0.73 | 0.46 | -4.14 | 3.43E-05 | **0.0001029** |

***Note: fold change is A-rule / V-rule***

**Figure 6 - source data 1.C**

**Analysis: Stimulus response modulation across rules by best frequency, BS cells**

| *Stimulus response* | *BF tuning* | *n* | *Mean FR resp across rules (fold change)* | *±SD* | *Kruskal-Wallis non-parametric ANOVA H-stat* | *p-value* | *FDR-corrected p-value* |
|---|---|---|---|---|---|---|---|
| AR* | Near AR | 44 | 0.96 | 0.71 | 3.39 | 0.184 | 0.184 |
| | Near AU | 10 | 1.35 | 1.81 | | | |
| | Other | 51 | 0.97 | 0.92 | | | |
| AU* | Near AR | 43 | 0.99 | 1.09 | 5.55 | 0.062 | 0.124 |
| | Near AU | 10 | 1.06 | 0.97 | | | |
| | Other | 52 | 0.78 | 0.85 | | | |

***Note: fold change is A-rule / V-rule***

**Figure 6 - source data 1.D**

**Analysis: Stimulus response modulation across rules by best frequency, NS cells**

| *Stimulus response* | *BF tuning* | *n* | *Mean FR resp across rules (fold change)* | *±SD* | *Kruskal-Wallis non-parametric ANOVA H-stat* | *p-value* | *FDR-corrected p-value* |
|---|---|---|---|---|---|---|---|
| AR* | Near AR | 24 | 0.94 | 0.44 | 0.53 | 0.768 | 0.768 |
| | Near AU | 7 | 1.04 | 0.10 | | | |
| | Other | 18 | 0.83 | 0.96 | | | |
| AU* | Near AR | 24 | 0.95 | 0.62 | 1.24 | 0.537 | 0.768 |
| | Near AU | 7 | 1.10 | 0.40 | | | |
| | Other | 18 | 0.76 | 0.96 | | | |

***Note: fold change is A-rule / V-rule***

**Figure 7 - source data 1.A**

**Analysis: decoder accuracy, A-rule vs V-rule by stim. type**

**Accuracy (%) ± SD**

| *Group* | ARVR | ARVU | AUVR | AUVU |
|---|---|---|---|---|
| BS | 56.5 ± 12.2 | 57.1 ± 11.6 | 57.4 ± 11.5 | 56.7 ± 12.3 |
| NS | 60.8 ± 13.4 | 62.1 ± 12.2 | 62.8 ± 13.1 | 60.0 ± 12.4 |

One-way Wilcoxon signed rank (WSR) vs chance [50%]: all p≤1.1E-12, all |z|≥7.1

**Figure 7 - source data 1.B**

**Analysis: decoder accuracy, A- vs V-rule by depth, BS/NS**

**Accuracy (%) ± SD**

| *Group* | Avg., all stim |
|---|---|
| Super. BS | 56.0 ± 10.8 |
| Super. NS | 59.6 ± 8.9 |
| Mid. BS | 55.8 ± 10.3 |
| Mid. NS | 60.5 ± 12.6 |
| Deep BS | 57.2 ± 10.5 |
| Deep NS | 62.9 ± 13.4 |

One-way WSR vs chance [50%]: all p≤2.9E-4, all |z|≥6.3

1153

**Figure 7 - source data 2.A**

**Analysis: decoder accuracy of AR* vs AU* across depth groups**

| *Group* | *n* | *A-rule % correct* | *±SD* | *WSR Z-value (vs 50% [chance])* | *p-value* | *V-rule % correct* | *±SD* | *WSR Z-value (vs 50% [chance])* | *p-value* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 48 | 62.58 | 12.66 | 5.18 | 2.20E-07 | 62.58 | 13 | 5.21 | 1.90E-07 |
| Super. NS | 30 | 71.14 | 15.49 | 4.51 | 6.30E-06 | 71.93 | 14.02 | 4.58 | 4.70E-06 |
| Mid. BS | 98 | 63.19 | 14.85 | 7.33 | 2.20E-13 | 63.54 | 14.79 | 7.57 | 3.60E-14 |
| Mid. NS | 25 | 73.96 | 13.62 | 4.35 | 1.40E-05 | 72.35 | 14.14 | 4.37 | 1.20E-05 |
| Deep BS | 233 | 67.81 | 15.19 | 12.5 | 7.20E-36 | 67.09 | 15.3 | 12.63 | 1.40E-36 |
| Deep NS | 61 | 72.18 | 15.74 | 6.45 | 1.10E-10 | 71.97 | 15.99 | 6.48 | 9.50E-11 |

**Figure 7 - source data 2.B**

**Analysis: decoder accuracy of ARVR vs AUVR across depth groups**

| *Group* | *n* | *A-rule % correct* | *±SD* | *WSR Z-value (vs 50% [chance])* | *p-value* | *V-rule % correct* | *±SD* | *WSR Z-value (vs 50% [chance])* | *p-value* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 45 | 63.54 | 13.4 | 4.93 | 8.30E-07 | 63.8 | 13.75 | 5.06 | 4.10E-07 |
| Super. NS | 30 | 70.67 | 16.23 | 4.45 | 8.50E-06 | 71.12 | 15.88 | 4.25 | 2.20E-05 |
| Mid. BS | 94 | 63.47 | 15.37 | 6.85 | 7.60E-12 | 63.01 | 16.36 | 6.34 | 2.30E-10 |
| Mid. NS | 25 | 73.35 | 14.39 | 4.37 | 1.20E-05 | 71.95 | 14.62 | 4.32 | 1.60E-05 |
| Deep BS | 207 | 69.03 | 15.86 | 11.62 | 3.40E-31 | 68.27 | 16.31 | 11.35 | 7.80E-30 |
| Deep NS | 60 | 71.99 | 15.39 | 6.42 | 1.40E-10 | 72.02 | 17.35 | 6.14 | 8.30E-10 |

**Figure 7 - source data 2.C**

**Analysis: decoder accuracy of ARVU vs AUVU across depth groups**

| Group | n | A-rule % correct | ±SD | WSR Z-value (vs 50% [chance]) | p-value | V-rule % correct | ±SD | WSR Z-value (vs 50% [chance]) | p-value |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 45 | 62.67 | 13.67 | 4.83 | 1.40E-06 | 62.91 | 13.79 | 4.99 | 5.90E-07 |
| Super. NS | 29 | 71.08 | 16.25 | 4.36 | 1.30E-05 | 72.88 | 13.79 | 4.59 | 4.30E-06 |
| Mid. BS | 93 | 63.76 | 14.97 | 7.16 | 8.10E-13 | 64.57 | 15.21 | 7.35 | 2.00E-13 |
| Mid. NS | 25 | 74.56 | 14 | 4.35 | 1.40E-05 | 72.75 | 15.14 | 4.32 | 1.60E-05 |
| Deep BS | 224 | 67.68 | 16.21 | 11.63 | 3.00E-31 | 67.24 | 16.39 | 11.77 | 5.60E-32 |
| Deep NS | 61 | 72.58 | 17.42 | 6.2 | 5.60E-10 | 72.3 | 16.22 | 6.48 | 9.00E-11 |

**Figure 7 - source data 3.A**

**Analysis: Wilcoxon signed-rank comparison of AR* vs AU* decoding across rules**

***** For raw data values, refer to Figure 7 - source data 2.B***

| *Group* | *n* | *WSR Z-value across rules* | *p-value* |
|---|---|---|---|
| Super. BS | 48 | -0.3 | 0.77 |
| Super. NS | 30 | 0.69 | 0.49 |
| Mid. BS | 98 | 0.61 | 0.54 |
| Mid. NS | 25 | -1.39 | 0.17 |
| Deep BS | 233 | -1.11 | 0.27 |
| Deep NS | 61 | -0.41 | 0.68 |

**Figure 7 - source data 3.B**

**Analysis: Wilcoxon signed-rank comparison of ARVR vs AUVR decoding accuracy across rules**

***** For raw data values, refer to Figure 7 - source data 2.B***

| *Group* | *n* | *WSR Z-value across rules* | *p-value* |
|---|---|---|---|
| Super. BS | 45 | 0.06 | 0.95 |
| Super. NS | 30 | 0.38 | 0.71 |
| Mid. BS | 94 | -0.17 | 0.87 |
| Mid. NS | 25 | -0.9 | 0.37 |
| Deep BS | 207 | -0.92 | 0.36 |
| Deep NS | 60 | 0.57 | 0.57 |

**Figure 7 - source data 3.C**

**Analysis: Wilcoxon signed-rank comparison of ARVU vs AUVU decoding accuracy across rules**

***** For raw data values, refer to Figure 7 - source data 2.C***

| *Group* | *n* | *WSR Z-value across rules* | *p-value* |
|---|---|---|---|
| Super. BS | 45 | 0.13 | 0.9 |
| Super. NS | 29 | 0.9 | 0.37 |
| Mid. BS | 93 | 0.81 | 0.42 |
| Mid. NS | 25 | -1.17 | 0.24 |
| Deep BS | 224 | -0.53 | 0.59 |
| Deep NS | 61 | -0.87 | 0.39 |

**Figure 7 - source data 4.A**

**Analysis: MI efficiency of ARVR vs AUVR across rules**

| *Group* | *n* | *A-rule mean bits/spk* | *±SD* | *V-rule mean bits/spk* | *±SD* | *WSR Z-value across rules* | *p-value* | *FDR-adjusted p-value* | *Median fold change across rules (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 48 | 0.18 | 0.14 | 0.15 | 0.11 | -1.47 | 0.14 | 0.396 | -- |
| Super. NS | 30 | 0.1 | 0.09 | 0.08 | 0.06 | -1.08 | 0.28 | 0.396 | -- |
| Mid. BS | 98 | 0.13 | 0.11 | 0.14 | 0.13 | 0.78 | 0.43 | 0.43 | -- |
| Mid. NS | 25 | 0.07 | 0.07 | 0.07 | 0.08 | -0.98 | 0.33 | 0.396 | -- |
| Deep BS | 233 | 0.19 | 0.19 | 0.15 | 0.13 | -4.06 | 4.90E-05 | 2.94E-04 | 1.19 |
| Deep NS | 61 | 0.06 | 0.06 | 0.06 | 0.05 | -1.28 | 0.2 | 0.396 | -- |

**Figure 7 - source data 4.B**

**Analysis: MI efficiency of ARVR vs AUVR across rules**

| *Group* | *n* | *A-rule mean bits/spk* | *±SD* | *V-rule mean bits/spk* | *±SD* | *WSR Z-value across rules* | *p-value* | *FDR-adjusted p-value* | *Median fold change across rules (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 45 | 0.16 | 0.13 | 0.14 | 0.14 | -1.31 | 0.19 | 0.57 | -- |
| Super. NS | 30 | 0.1 | 0.09 | 0.08 | 0.07 | -1 | 0.32 | 0.64 | -- |
| Mid. BS | 94 | 0.13 | 0.13 | 0.14 | 0.13 | 0.62 | 0.54 | 0.81 | -- |
| Mid. NS | 25 | 0.07 | 0.08 | 0.07 | 0.08 | -0.31 | 0.76 | 0.91 | -- |
| Deep BS | 207 | 0.17 | 0.18 | 0.13 | 0.12 | -3.9 | 9.70E-05 | 5.82E-04 | 1.2 |
| Deep NS | 60 | 0.06 | 0.06 | 0.06 | 0.06 | -0.12 | 0.91 | 0.91 | -- |

**Figure 7 - source data 4.C**

**Analysis: MI efficiency of ARVU vs AUVU across rules**

| *Group* | *n* | *A-rule mean bits/spk* | *±SD* | *V-rule mean bits/spk* | *±SD* | *WSR Z-value across rules* | *p-value* | *FDR-adjusted p-value* | *Median fold change across rules (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 45 | 0.18 | 0.18 | 0.15 | 0.12 | -0.34 | 0.73 | 0.8 | -- |
| Super. NS | 29 | 0.1 | 0.1 | 0.07 | 0.06 | -0.25 | 0.8 | 0.8 | -- |
| Mid. BS | 93 | 0.11 | 0.1 | 0.13 | 0.13 | 0.66 | 0.51 | 0.765 | -- |
| Mid. NS | 25 | 0.08 | 0.07 | 0.07 | 0.08 | -0.79 | 0.43 | 0.765 | -- |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Deep BS | 224 | 0.18 | 0.18 | 0.15 | 0.14 | -2.75 | 0.006 | 0.036 | 1.15 |
| Deep NS | 61 | 0.06 | 0.06 | 0.06 | 0.05 | -1.33 | 0.19 | 0.57 | -- |

**Figure 7 - source data 5.**

**Analysis: Comparison of STRF MI (bits/spk) across rules**

| *Group* | *n* | *A-rule mean bits/spk* | *±SD* | *V-rule mean bits/spk* | *±SD* | *WSR Z-value across rules* | *p-value* | *FDR-adjusted p-value* | *Median fold change across rules (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| Super. BS | 8 | 0.16 | 0.21 | 0.16 | 0.22 | -0.17 | 0.87 | 0.92 | -- |
| Super. NS | 8 | 0.07 | 0.08 | 0.07 | 0.08 | -1.54 | 0.12 | 0.24 | -- |
| Mid. BS | 27 | 0.08 | 0.1 | 0.06 | 0.07 | -0.88 | 0.38 | 0.57 | -- |
| Mid. NS | 13 | 0.03 | 0.03 | 0.04 | 0.05 | 0.1 | 0.92 | 0.92 | -- |
| Deep BS | 50 | 0.15 | 0.22 | 0.13 | 0.21 | -3.05 | 0.0023 | 0.0138 | 1.25 |
| Deep NS | 14 | 0.04 | 0.05 | 0.03 | 0.03 | -1.66 | 0.096 | 0.24 | -- |

**Figure 8 - source data 1.A**

**Analysis: Pre-stimulus FRs for false alarm vs. correct reject in A-rule**

**Inclusion filter: behavior sessions with >10 FA and CR trials in A-rule**

| Group | *n* | *CR trials mean pre-stim FR (Hz)* | *±SD* | *FA trials mean pre-stim FR (Hz)* | *±SD* | *WSR Z-value across outcomes* | *p-val* | *FDR-adjusted p-val* | *Mean FR difference (Hz): FA-CR (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| | ***Units with enhanced FRs in A-rule*** | | | | | | | | |
| Super. BS | 10 | 9.47 | 9.71 | 8.52 | 9.51 | 1.17 | 0.2411 | 0.2893 | -- |
| Mid. BS | 18 | 3.66 | 3.89 | 3.09 | 2.65 | 0.46 | 0.6474 | 0.6474 | -- |
| Deep BS | 63 | 3.61 | 4.43 | 3.51 | 4.89 | 1.43 | 0.1515 | 0.2893 | -- |
| | ***Units with suppressed FRs in A-rule*** | | | | | | | | |
| Super. BS | 9 | 2.14 | 2.63 | 2.51 | 2.53 | -1.48 | 0.1386 | 0.2893 | -- |
| Mid. BS | 36 | 3.12 | 3.69 | 3.59 | 3.73 | -1.24 | 0.2146 | 0.2893 | -- |
| Deep BS | 98 | 3.36 | 3.80 | 3.71 | 3.64 | -3.15 | 0.0016 | 0.0098 | 0.35 |

**Figure 8 - source data 1.B**

**Analysis: Pre-stimulus FRs for false alarm vs. correct reject in V-rule**

**Inclusion filter: behavior recordings with >10 FA and CR trials in V-rule**

| *Group* | *n (units)* | *CR trials mean pre-stim FR (Hz)* | *±SD* | *FA trials mean pre-stim FR (Hz)* | *±SD* | *WSR Z-value across outcomes* | *p-val* | *FDR-adjusted p-val* | *Mean FR difference (Hz): FA-CR (FDR-adj p<0.05 only)* |
|---|---|---|---|---|---|---|---|---|---|
| | ***Units with enhanced FRs in A-rule*** | | | | | | | | |
| Super. BS | 10 | 5.91 | 5.49 | 5.34 | 4.57 | 0.36 | 0.7213 | 0.7369 | -- |
| Mid. BS | 17 | 7.66 | 8.25 | 8.00 | 9.22 | -0.54 | 0.5862 | 0.7369 | -- |
| Deep BS | 27 | 3.11 | 3.44 | 3.71 | 4.51 | -1.59 | 0.1128 | 0.6769 | -- |
| | ***Units with suppressed FRs in A-rule*** | | | | | | | | |
| Super. BS | 4 | 4.03 | 3.66 | 4.92 | 5.38 | -0.73 | 0.4652 | 0.7369 | -- |
| Mid. BS | 27 | 5.38 | 4.84 | 5.21 | 5.91 | 0.86 | 0.3871 | 0.7369 | -- |
| Deep BS | 95 | 5.72 | 7.15 | 5.67 | 7.61 | 0.34 | 0.7369 | 0.7369 | -- |

# References


Aertsen AM, Johannesma PI. 1981. The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol Cybern* **42**:133–143.

Arieli A, Sterkin A, Grinvald A, Aertsen A. 1996. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* **273**:1868–1871.

Atencio CA, Schreiner CE. 2013. Stimulus choices for spike-triggered receptive field analysis. *Handbook of modern techniques in auditory cortex New York: Nova Biomedical* 61–100.

Atencio CA, Schreiner CE. 2012. Spectrotemporal processing in spectral tuning modules of cat primary auditory cortex. *PLoS One* **7**:e31537.

Atencio CA, Sharpee TO, Schreiner CE. 2008. Cooperative nonlinearities in auditory cortical neurons. *Neuron* **58**:956–966.

Atiani S, Elhilali M, David SV, Fritz J, Shamma S. 2009. Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron* **61**:467–480.

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.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. doi:10.1111/j.2517-6161.1995.tb02031.x

Bigelow J, Morrill RJ, Dekloe J, Hasenstaub AR. 2019. Movement and VIP interneuron activation differentially modulate encoding in mouse auditory cortex. *Eneuro* **6**.

Bigelow J, Morrill RJ, Olsen T, Hasenstaub AR. 2022. Visual modulation of firing and spectrotemporal receptive fields in mouse auditory cortex. *Current Research in Neurobiology* **3**:100040.

Bisley JW, Goldberg ME. 2003. Neuronal activity in the lateral intraparietal area and spatial attention. *Science* **299**:81–86.

Bradley MM, Miccoli L, Escrig MA, Lang PJ. 2008. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **45**:602–607.

Budinger E, Laszcz A, Lison H, Scheich H, Ohl FW. 2008. Non-sensory cortical and subcortical connections of the primary auditory cortex in Mongolian gerbils: bottom-up and top-down processing of neuronal information via field AI. *Brain Res* **1220**:2–32.

Budinger E, Scheich H. 2009. Anatomical connections suitable for the direct processing of neuronal information of different modalities via the rodent primary auditory cortex. *Hear Res* **258**:16–27.

Buracas GT, Zador AM, DeWeese MR, Albright TD. 1998. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron* **20**:959–969.

Buran BN, von Trapp G, Sanes DH. 2014. Behaviorally gated reduction of spontaneous discharge can improve detection thresholds in auditory cortex. *J Neurosci* **34**:4076–4081.

Carcea I, Insanally MN, Froemke RC. 2017. Dynamics of auditory cortical activity during behavioural engagement and auditory perception. *Nat Commun* **8**:14412.

Cardin JA, Palmer LA, Contreras D. 2007. Stimulus feature selectivity in excitatory and inhibitory neurons in primary visual cortex. *J Neurosci* **27**:10333–10344.

Christianson GB, Sahani M, Linden JF. 2011. Depth-dependent temporal response properties in core auditory cortex. *J Neurosci* **31**:12837–12848.

Cohen MR, Maunsell JHR. 2009. Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* **12**:1594–1600.

Cools R, Clark L, Robbins TW. 2004. Differential responses in human striatum and prefrontal cortex to changes in object and rule relevance. *J Neurosci* **24**:1129–1135.

Cox MA, Dougherty K, Adams GK, Reavis EA, Westerberg JA, Moore BS, Leopold DA, Maier A. 2019. Spiking Suppression Precedes Cued Attentional Enhancement of Neural Responses in Primary Visual Cortex. *Cereb Cortex* **29**:77–90.

Crone EA, Wendelken C, Donohue SE, Bunge SA. 2006. Neural evidence for dissociable components of task-switching. *Cereb Cortex* **16**:475–486.

Da Costa S, van der Zwaag W, Miller LM, Clarke S, Saenz M. 2013. Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. *J Neurosci* **33**:1858–1863.

de Boer E. 1968. Reverse Correlation I. - A heuristic introduction to the technique of triggered correlation with application to the analysis of compound systems. *Proc Kon Ned Acad Wetensch* **71**:472–486.

Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Annu Rev Neurosci* **18**:193–222.

DiCarlo JJ, Lane JW, Hsiao SS, Johnson KO. 1996. Marking microelectrode penetrations with fluorescent dyes. *J Neurosci Methods* **64**:75–81.

Downer JD, Niwa M, Sutter ML. 2015. Task engagement selectively modulates neural correlations in primary auditory cortex. *J Neurosci* **35**:7565–7574.

Egner T, Monti JM, Summerfield C. 2010. Expectation and surprise determine neural population responses in the ventral visual stream. *J Neurosci* **30**:16601–16608.

Escabí MA, Read HL, Viventi J, Kim D-H, Higgins NC, Storace DA, Liu ASK, Gifford AM, Burke JF, Campisi M, Kim Y-S, Avrin AE, Spiegel Jan V der, Huang Y, Li M, Wu J, Rogers JA, Litt B, Cohen YE. 2014. A high-density, high-channel count, multiplexed μECoG array for auditory-cortex recordings. *J Neurophysiol* **112**:1566–1583.

Escab□MA, Schreiner CE. 2002. Nonlinear Spectrotemporal Sound Analysis by Neurons in the Auditory Midbrain. *J Neurosci* **22**:4114–4131.

Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* **1**:1–47.

Fiáth R, Márton AL, Mátyás F, Pinke D, Márton G, Tóth K, Ulbert I. 2019. Slow insertion of silicon probes improves the quality of acute neuronal recordings. *Sci Rep* **9**:111.

Foffani G, Moxon KA. 2004. PSTH-based classification of sensory stimuli using ensembles of single neurons. *J Neurosci Methods* **135**:107–120.

Fritz J, Elhilali M, David SV, Shamma S. 2007. Auditory attention—focusing the searchlight on sound. *Curr Opin Neurobiol* **17**:437–455.

Fritz J, Elhilali M, Shamma S. 2005. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *Journal of Neuroscience*.

Fritz J, Shamma S, Elhilali M, Klein D. 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* **6**:1216–1223.

Gourévitch B, Occelli F, Gaucher Q, Aushana Y, Edeline J-M. 2015. A new and fast characterization of multiple encoding properties of auditory neurons. *Brain Topogr* **28**:379–400.

Guo ZV, Hires SA, Li N, O'Connor DH, Komiyama T, Ophir E, Huber D, Bonardi C, Morandell K, Gutnisky D, Peron S, Xu N-L, Cox J, Svoboda K. 2014. Procedures for behavioral experiments in head-fixed mice. *PLoS One* **9**:e88678.

Haider B, McCormick DA. 2009. Rapid neocortical dynamics: cellular and network mechanisms. *Neuron* **62**:171–189.

Harris KD, Thiele A. 2011. Cortical state and attention. *Nat Rev Neurosci* **12**:509–523.

Hautus MJ. 1995. Corrections for extreme proportions and their biasing effects on estimated values ofd′. *Behav Res Methods Instrum Comput* **27**:46–51.

Herrington TM, Assad JA. 2010. Temporal sequence of attentional modulation in the lateral intraparietal area and middle temporal area during rapid covert shifts of attention. *J Neurosci* **30**:3287–3296.

Hess EH, Polt JM. 1964. Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science* **143**:1190–1192.
Hocherman S, Benson DA, Goldstein MH Jr, Heffner HE, Hienz RD. 1976. Evoked unit activity in auditory cortex of monkeys performing a selective attention task. *Brain Res* **117**:51–68.
Hoglen NEG, Larimer P, Phillips EAK, Malone BJ, Hasenstaub AR. 2018. Amplitude modulation coding in awake mice and squirrel monkeys. *J Neurophysiol* **119**:1753–1766.
Johnson JA, Zatorre RJ. 2005. Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. *Cereb Cortex* **15**:1609–1620.
Kawaguchi K, Clery S, Pourriahi P, Seillier L, Haefner RM, Nienborg H. 2018. Differentiating between models of perceptual decision making using pupil size inferred confidence. *J Neurosci* **38**:8874–8888.
Kleiner M, Brainard D, Pelli D. 2007. What's new in Psychtoolbox-3?
Krishna BS, Semple MN. 2000. Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol* **84**:255–273.
Kuchibhotla KV, Gill JV, Lindsay GW, Papadoyannis ES, Field RE, Sten TAH, Miller KD, Froemke RC. 2017. Parallel processing by cortical inhibition enables context-dependent behavior. *Nat Neurosci* **20**:62–71.
Laboy-Juárez KJ, Ahn S, Feldman DE. 2019. A normalized template matching method for improving spike detection in extracellular voltage recordings. *Sci Rep* **9**:12087.
Lakatos P, Barczak A, Neymotin SA, McGinnis T, Ross D, Javitt DC, O'Connell MN. 2016. Global dynamics of selective attention and its lapses in primary auditory cortex. *Nat Neurosci* **19**:1707–1717.
Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE. 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* **320**:110–113.
Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE. 2009. The leading sense: supramodal control of neurophysiological context by attention. *Neuron* **64**:419–430.
Lamme VA, Supèr H, Spekreijse H. 1998. Feedforward, horizontal, and feedback processing in the visual cortex. *Curr Opin Neurobiol* **8**:529–535.
Land R, Engler G, Kral A, Engel AK. 2013. Response properties of local field potentials and multiunit activity in the mouse visual cortex. *Neuroscience* **254**:141–151.
Li L-Y, Xiong XR, Ibrahim LA, Yuan W, Tao HW, Zhang LI. 2015. Differential Receptive Field Properties of Parvalbumin and Somatostatin Inhibitory Neurons in Mouse Auditory Cortex. *Cereb Cortex* **25**:1782–1791.
Licata AM, Kaufman MT, Raposo D, Ryan MB, Sheppard JP, Churchland AK. 2017. Posterior Parietal Cortex Guides Visual Decisions in Rats. *J Neurosci* **37**:4954–4966.
Lima SQ, Hromádka T, Znamenskiy P, Zador AM. 2009. PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. *PLoS One* **4**:e6099.
Malone BJ, Scott BH, Semple MN. 2007. Dynamic amplitude coding in the auditory cortex of awake rhesus macaques. *J Neurophysiol* **98**:1451–1474.
Maunsell JHR, Treue S. 2006. Feature-based attention in visual cortex. *Trends Neurosci* **29**:317–322.
McGinley MJ, David SV, McCormick DA. 2015. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron* **87**:179–192.
Mitchell JF, Sundberg KA, Reynolds JH. 2007. Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* **55**:131–141.
Moran J, Desimone R. 1985. Selective attention gates visual processing in the extrastriate cortex. *Science* **229**:782–784.
Morrill RJ, Hasenstaub AR. 2018. Visual Information Present in Infragranular Layers of Mouse

Auditory Cortex. *J Neurosci* **38**:2854–2862.
Morrill RJ, Bigelow J, DeKloe J, Hasenstaub AR 2022. AVtrainer-stim. Github. https://github.com/HasenstaubLab/AVtrainer-stim/tree/main/demo 737720F
Nandy AS, Nassi JJ, Reynolds JH. 2017. Laminar Organization of Attentional Modulation in Macaque Visual Area V4. *Neuron* **93**:235–246.
O'Connell MN, Barczak A, Schroeder CE, Lakatos P. 2014. Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. *J Neurosci* **34**:16496–16508.
Otazu GH, Tai L-H, Yang Y, Zador AM. 2009. Engaging in an auditory task suppresses responses in auditory cortex. *Nat Neurosci* **12**:646–654.
Pachitariu M, Steinmetz N, Kadir S, Carandini M, D HK. 2016. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *Cold Spring Harbor Laboratory*. doi:10.1101/061481
Park H, Ince RAA, Schyns PG, Thut G, Gross J. 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr Biol* **25**:1649–1653.
Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL. 2004. Attentional modulation of human auditory cortex. *Nat Neurosci* **7**:658–663.
Phillips EAK, Hasenstaub AR. 2016. Asymmetric effects of activating and inactivating cortical interneurons. *Elife* **5**. doi:10.7554/eLife.18383
Phillips EAK, Schreiner CE, Hasenstaub AR. 2017. Diverse effects of stimulus history in waking mouse auditory cortex. *J Neurophysiol* **118**:1376–1393.
Reimer J, Froudarakis E, Cadwell CR, Yatsenko D, Denfield GH, Tolias AS. 2014. Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* **84**:355–362.
Reynolds JH, Chelazzi L. 2004. Attentional modulation of visual processing. *Annu Rev Neurosci* **27**:611–647.
Rikhye RV, Gilra A, Halassa MM. 2018. Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nat Neurosci* **21**:1753–1763.
Rodgers CC, DeWeese MR. 2014. Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task for rodents. *Neuron* **82**:1157–1170.
Roedel A, Storch C, Holsboer F, Ohl F. 2006. Effects of light or dark phase testing on behavioural and cognitive performance in DBA mice. *Lab Anim* **40**:371–381.
Rougier NP, Noelle DC, Braver TS, Cohen JD, O'Reilly RC. 2005. Prefrontal cortex and flexible cognitive control: rules without symbols. *Proc Natl Acad Sci U S A* **102**:7338–7343.
Rutkowski RG, Shackleton TM, Schnupp JWH, Wallace MN, Palmer AR. 2002. Spectrotemporal receptive field properties of single units in the primary, dorsocaudal and ventrorostral auditory cortex of the guinea pig. *Audiol Neurootol* **7**:214–227.
Saderi D, Schwartz ZP, Heller CR, Pennington JR, David SV. 2021. Dissociation of task engagement and arousal effects in auditory cortex and midbrain. *Elife* **10**. doi:10.7554/eLife.60153
Sakata S, Harris KD. 2009. Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron* **64**:404–418.
Salin PA, Bullier J. 1995. Corticocortical connections in the visual system: structure and function. *Physiol Rev* **75**:107–154.
Samuelsen CL, Gardner MPH, Fontanini A. 2012. Effects of cue-triggered expectation on cortical processing of taste. *Neuron* **74**:410–422.
Sato T, Schall JD. 2001. Pre-excitatory pause in frontal eye field responses. *Exp Brain Res* **139**:53–58.
Schriver BJ, Bagdasarov S, Wang Q. 2018. Pupil-linked arousal modulates behavior in rats

performing a whisker deflection direction discrimination task. *J Neurophysiol* **120**:1655–1670.
Senzai Y, Fernandez-Ruiz A, Buzsáki G. 2019. Layer-Specific Physiological Features and Interlaminar Interactions in the Primary Visual Cortex of the Mouse. *Neuron* **101**:500-513.e5.
Shomstein S, Yantis S. 2004. Control of attention shifts between vision and audition in human cortex. *J Neurosci* **24**:10702–10706.
Sukiban J, Voges N, Dembek TA, Pauli R, Visser-Vandewalle V, Denker M, Weber I, Timmermann L, Grün S. 2019. Evaluation of Spike Sorting Algorithms: Application to Human Subthalamic Nucleus Recordings and Simulations. *Neuroscience* **414**:168–185.
Sutter ML, Shamma SA. 2011. The relationship of auditory cortical activity to perception and behaviorThe Auditory Cortex. Boston, MA: Springer US. pp. 617–641.
Toth LJ, Assad JA. 2002. Dynamic coding of behaviourally relevant stimuli in parietal cortex. *Nature* **415**:165–168.
Treue S, Martínez Trujillo JC. 1999. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**:575–579.
Wimmer RD, Schmitt LI, Davidson TJ, Nakajima M, Deisseroth K, Halassa MM. 2015. Thalamic control of sensory selection in divided attention. *Nature* **526**:705–709.
Winkowski DE, Bandyopadhyay S, Shamma SA, Kanold PO. 2013. Frontal cortex activation causes rapid plasticity of auditory cortical processing. *J Neurosci* **33**:18134–18148.
Woodruff PW, Benson RR, Bandettini PA, Kwong KK, Howard RJ, Talavage T, Belliveau J, Rosen BR. 1996. Modulation of auditory and visual cortex by selective attention is modality-dependent. *Neuroreport* **7**:1909–1913.
Yin P, Fritz J, Shamma S. 2014. Rapid Spectrotemporal Plasticity in Primary Auditory Cortex during Behavior. *J Neurosci* **34**:4396–4408.
Yoshida T, Katz DB. 2011. Control of prestimulus activity related to improved sensory coding within a discrimination task. *J Neurosci* **31**:4101–4112.
Zador A. 1998. Impact of synaptic unreliability on the information transmitted by spiking neurons. *J Neurophysiol* **79**:1219–1229.
Zempeltzi MM, Kisse M, Brunk MGK, Glemser C, Aksit S, Deane KE, Maurya S, Schneider L, Ohl FW, Deliano M, Happel MFK. 2020. Task rule and choice are reflected by layer-specific processing in rodent auditory cortical microcircuits. *Commun Biol* **3**:345.