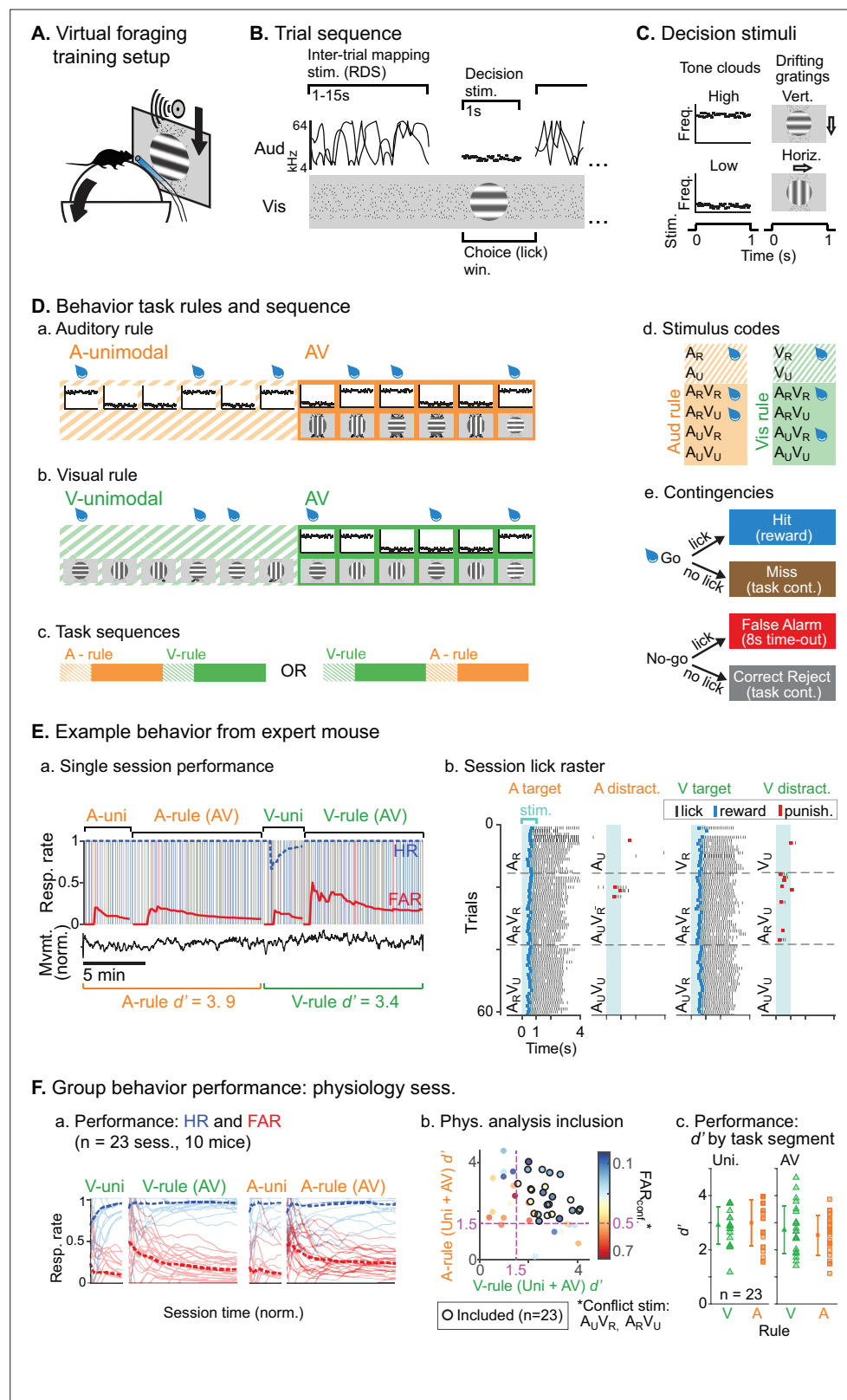


---

## Figures and figure supplements

Audiovisual task switching rapidly modulates sound encoding in mouse auditory cortex

**Ryan J Morrill *et al***

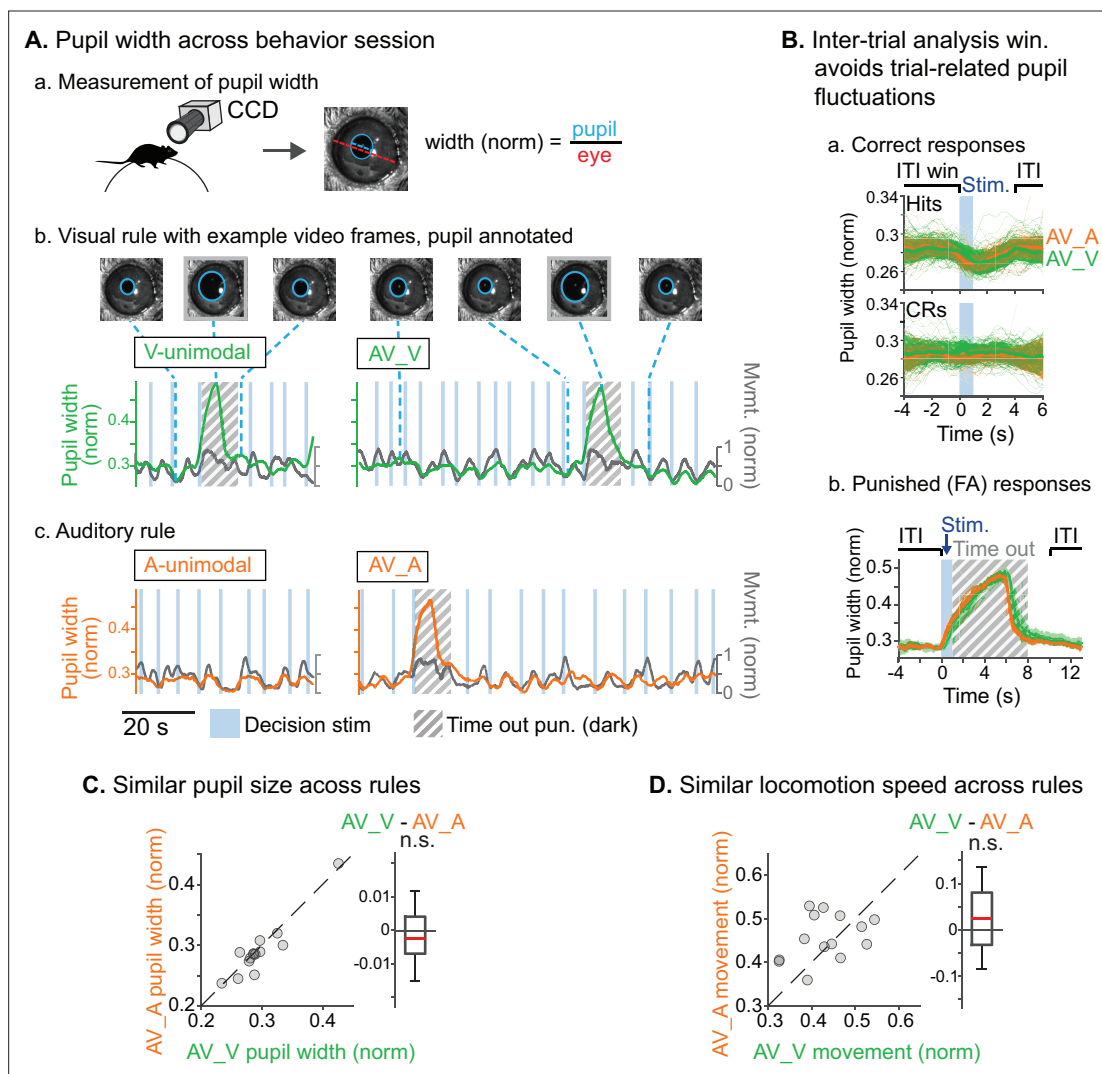


**Figure 1.** Audiovisual rule-switching in mice. **(A)** Virtual foraging environment: a head-fixed mouse runs on a floating spherical treadmill. Locomotion measured by treadmill movement controls auditory and visual stimulus presentation. A water spout in front of the mouse provides rewards. A lickometer records licks, which determines reward or punishment. **(B)** Trial sequence: during inter-trial intervals, a track of moving dots provides visual

Figure 1 continued on next page

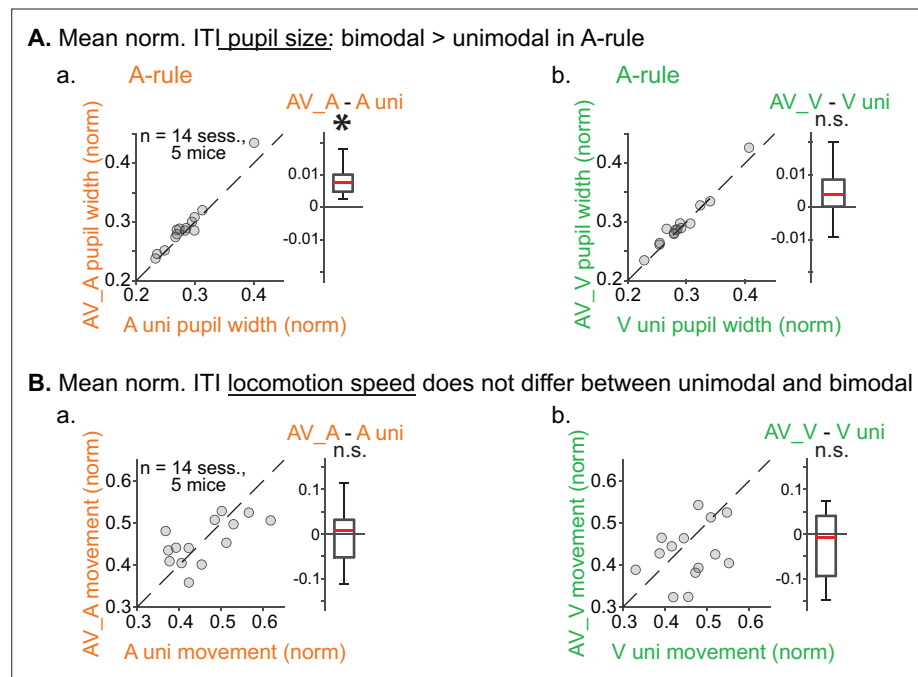
# Figure 1 continued

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. The 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 are tone clouds (TCs; 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 used one of two possible task sequences. **(d)** Stimulus codes, for reference. **(e)** Contingencies for water reward, timeout punishment, or task continuation. **(F)** 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:  $A_U V_R$  in A-rule or  $A_R V_U$  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.  $FAR_{conf} < 0.5$  for conflict stimuli, as a critical test of modality-selective attention. **(c)**  $d'$  is similar across task rules in unimodal and AV segments.

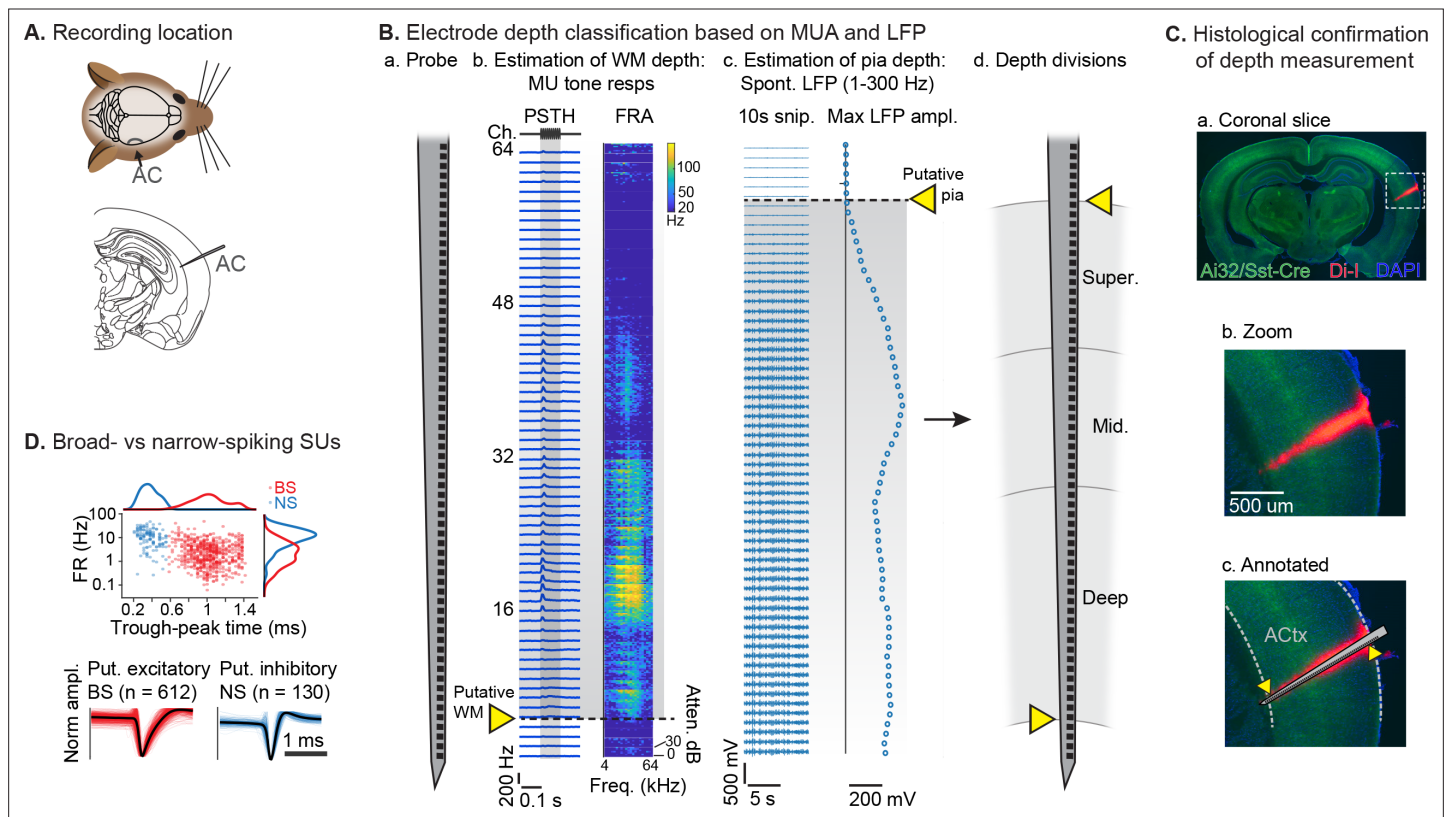


**Figure 2.** Similar levels of arousal and movement during auditory and visual attention. **(A)** Pupil size measurement. **(a)** Left eye pupil recorded via CCD camera during the task. Pupil circumference (light blue) is tracked using automated video analysis; size is measured as pupil diameter over 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 timeout 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. Difference box plots: central line: median; box edges: 25th and 75th percentiles; whiskers: data points not considered outliers. **(D)** Min-max-normalized locomotion is also similar across rules.

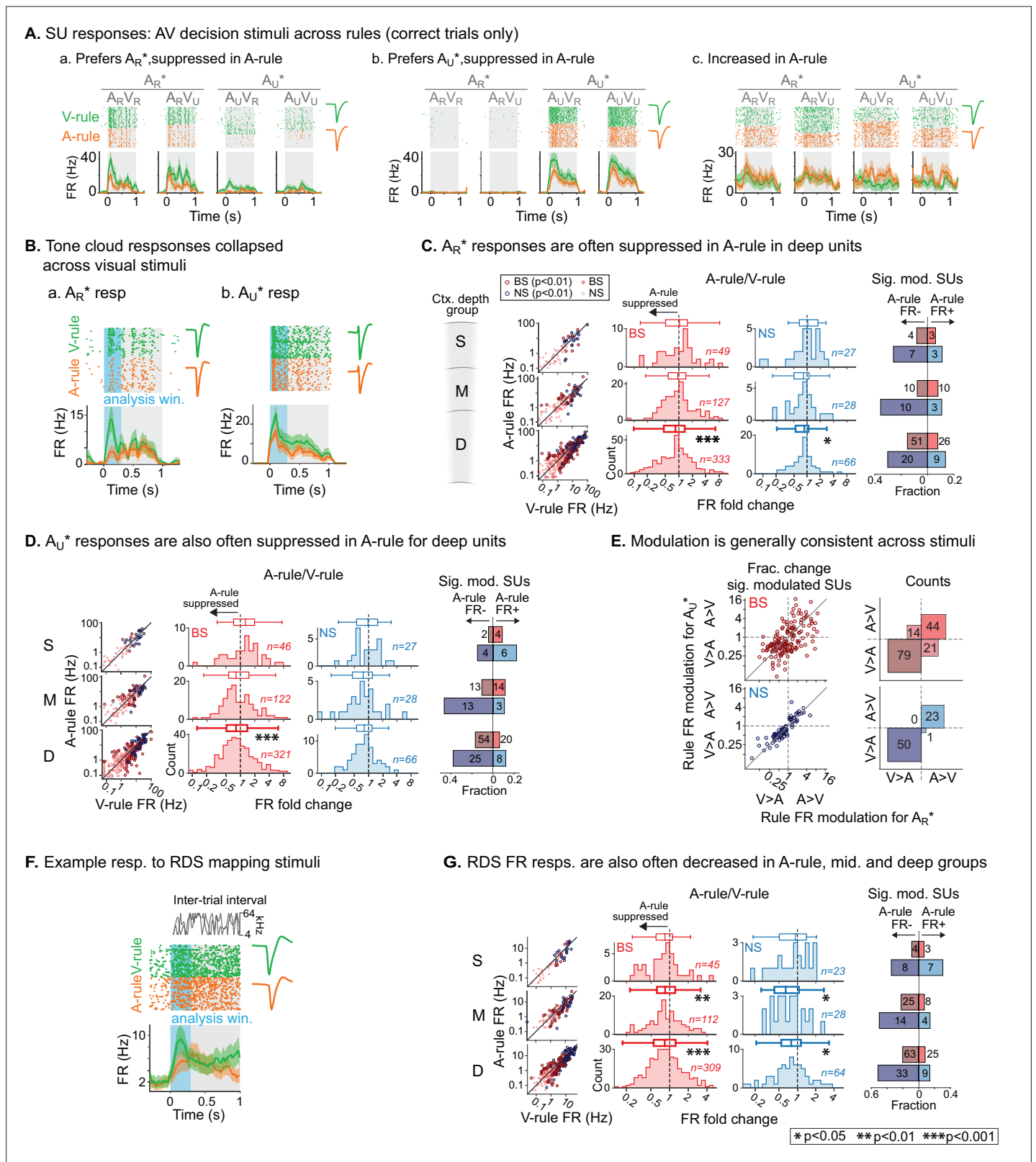




**Figure 2—figure supplement 1.** Pupil size and locomotion compared between unimodal and bimodal blocks. **(A)** Pupil size, measured as described in **Figure 2**, compared within rule between unimodal and bimodal task segments. **(a)** A-rule: unimodal A-only pupil size is smaller than bimodal AV pupil (A-rule unimodal:  $0.28 \pm 0.04$  mean norm, pupil width  $\pm$  SD, A-rule bimodal:  $0.29 \pm 0.05$ ;  $Z = -2.6$ ,  $p = 0.009$ ), 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:  $0.29 \pm 0.04$ , V-rule bimodal:  $0.30 \pm 0.05$ ;  $Z = -2.0$ ,  $p = 0.062$ ). **(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 unimodal:  $0.46 \pm 0.08$ , A bimodal:  $0.46 \pm 0.05$ ;  $Z = 0.282$ ,  $p = 0.78$ ). **(b)** V-rule: locomotion during unimodal is comparable to locomotion during bimodal (V uni:  $0.46 \pm 0.06$ , V bimodal:  $0.43 \pm 0.07$ ;  $Z = 1.224$ ,  $p = 0.33$ ). All p-values false discovery rate (FDR) adjusted for  $n = 3$  comparisons (including A-rule bimodal vs. V-rule bimodal in **Figure 2**). Conventions for difference box plots as in **Figure 2**.



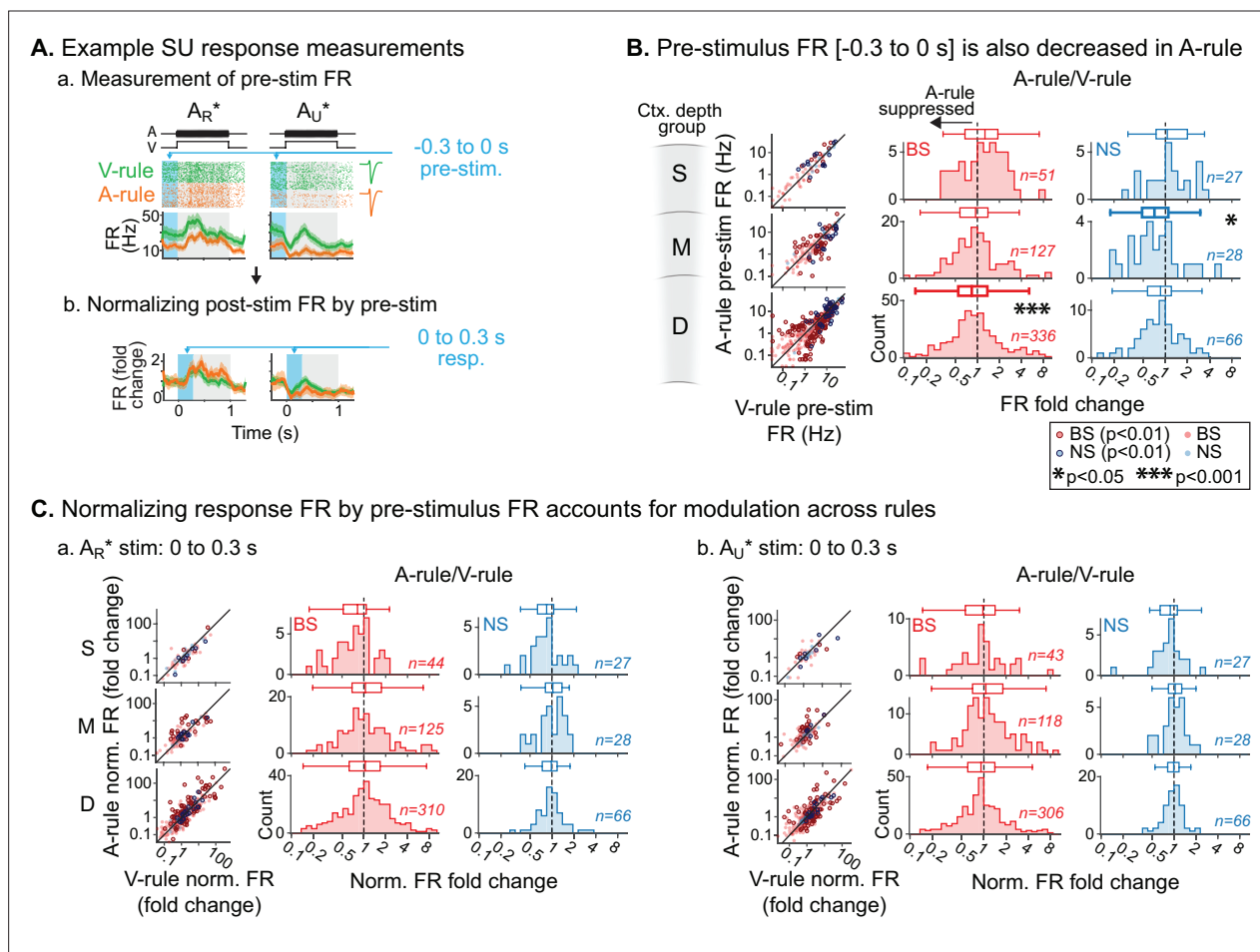
**Figure 3.** Single unit (SU) recording and depth estimation in auditory cortex. **(A)** Translaminar probes were used to record activity in right auditory cortex (AC). **(B)** Physiological estimation of cortical depth. **(a)** Linear 64-channel probe captures all activity in layers of AC. **(b)** Example tone-evoked multi-unit (MU) sound responses by channel, providing a marker for the border of deep cortex and white matter (WM). Left: peristimulus time histogram (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.



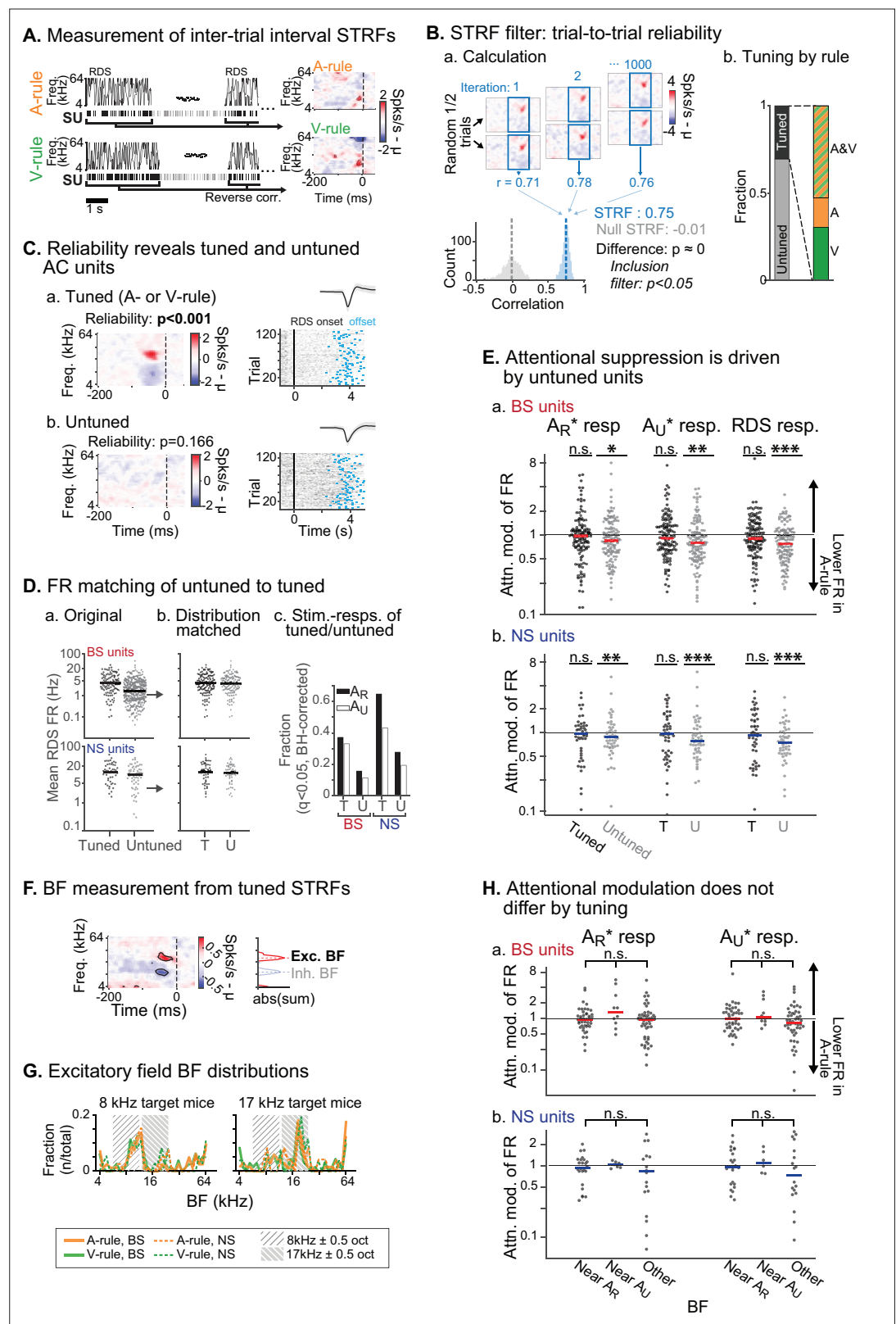
**Figure 4.** Net suppression of sound-evoked firing rates during auditory attention. **(A)** Example single unit (SU) responses to physically identical audiovisual (AV) stimuli across task rules ( $A_R^* = A_R V_R$  and  $A_R V_U$  collapsed;  $A_U^* = A_U V_U$  and  $A_U V_R$  collapsed). **(a)** Response showing 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 firing rate (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 Figure 4 continued on next page

## Figure 4 continued

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: broad-spiking (BS) units. Blue: narrow-spiking (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 firing rate (FR) measurement, and normalization of post-stimulus response. **(a)** Raw FR by condition and stimulus. Pre-stimulus analysis window shown in blue (−0.3–0 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 broad-spiking/narrow-spiking (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 false discovery rate (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.



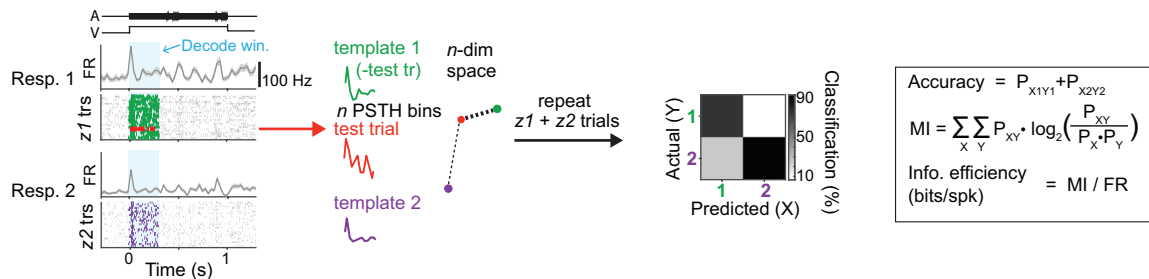
**Figure 6.** Attentional modulation of spike rate is driven by neurons without spectrotemporal receptive field (STRF) tuning. (A) STRFs for A-rule and V-rule were calculated from spikes during the inter-trial-interval random-double sweep (RDS) stimulus using standard reverse correlation methods. (B) STRF reliability as a measure for tuning. (a) Reliability was measured through correlations of randomly subsampled halves of all RDS presentations, Figure 6 continued on next page

# Figure 6 continued

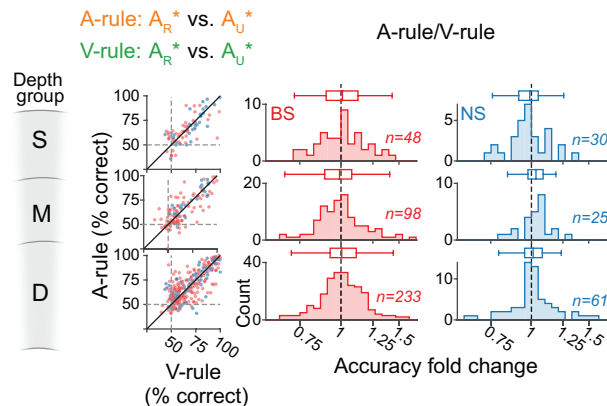
repeated 1000 times. A p-value was calculated empirically through comparison of correlation value distributions from the actual STRF and a null STRF, generated from random circular shuffling of spike trains relative to stimulus. **(b)** Left: fraction of tuned and untuned units. Right: fractions of units 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 broad-spiking [BS], narrow-spiking [NS]) is matched for sample size and firing rate to the group with tuned STRFs ( $n=121$ , 51 for BS, NS). **(a)** Mean firing rate (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  $A_U^*$  tone clouds [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 false discovery rate (FDR)-adjusted,  $q=0.01$ . **(e)** Untuned unit group is suppressed during auditory attention, while tuned unit group is not. **(a)** Attentional modulation of BS 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)** NS, 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  $A_R$  ( $BF \pm 0.5$  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).



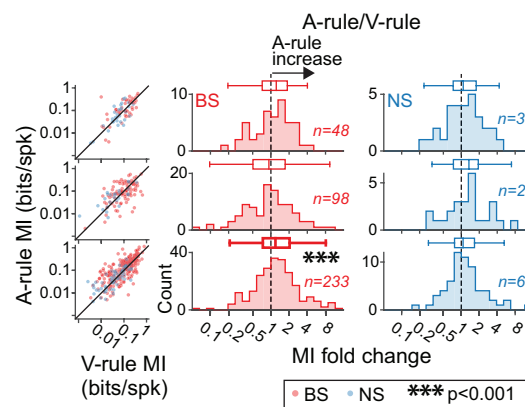
## A. Decision stim: PSTH-based Euclidean distance decoding



## B. Auditory stimulus decoding accuracy is similar across rules

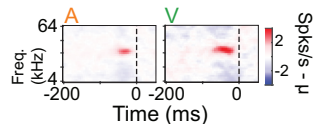


## C. Stimulus-spike information efficiency increases during A-rule in deep BS units

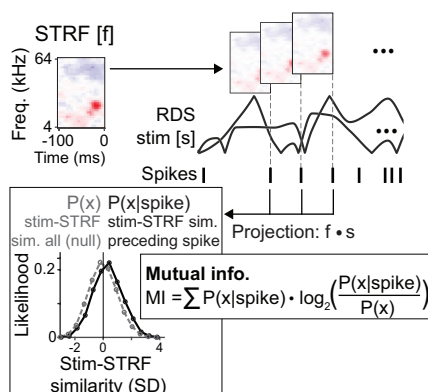


## D. ITI RDS stim: MI between spiking and linear STRF filter

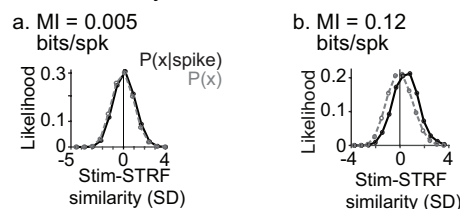
## a. STRFs calculated from A-rule and V-rule



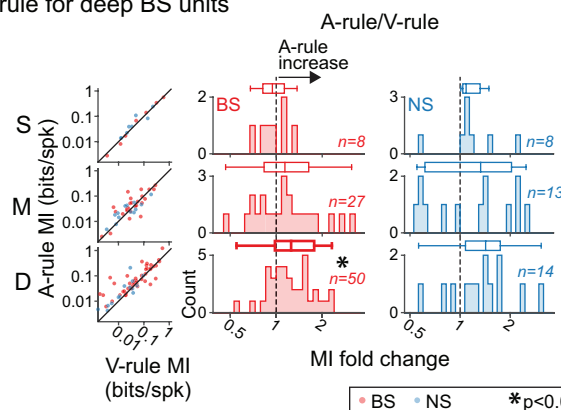
## b. Calculation of MI (bits/spk)



## E. Example MI efficiency



## F. ITI spiketrain-STRF information efficiency increases during A-rule for deep BS units



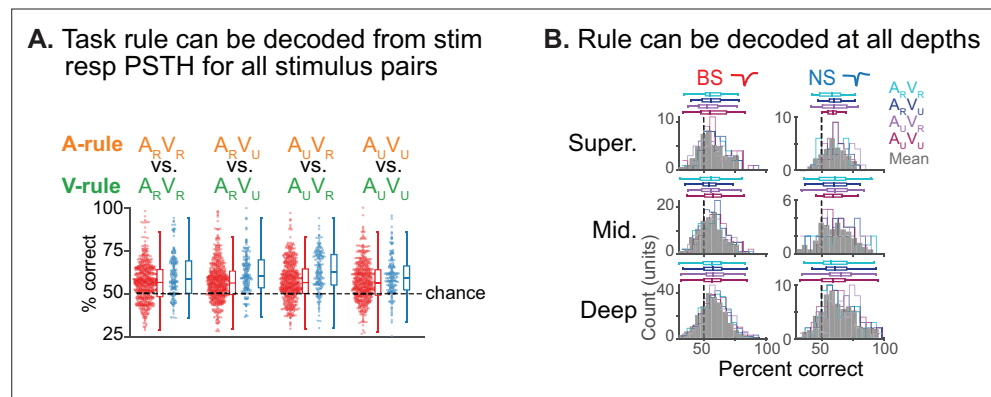
**Figure 7.** Auditory attention increases sound encoding efficiency in deep-layer broad-spiking units. (A) Peristimulus time histogram (PSTH)-based spike train decoding analysis. Time-binned responses for each single trial (test trial; red) are compared to PSTHs (templates; green, purple) reflecting 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, MI, and Info. efficiency.

Figure 7 continued on next page

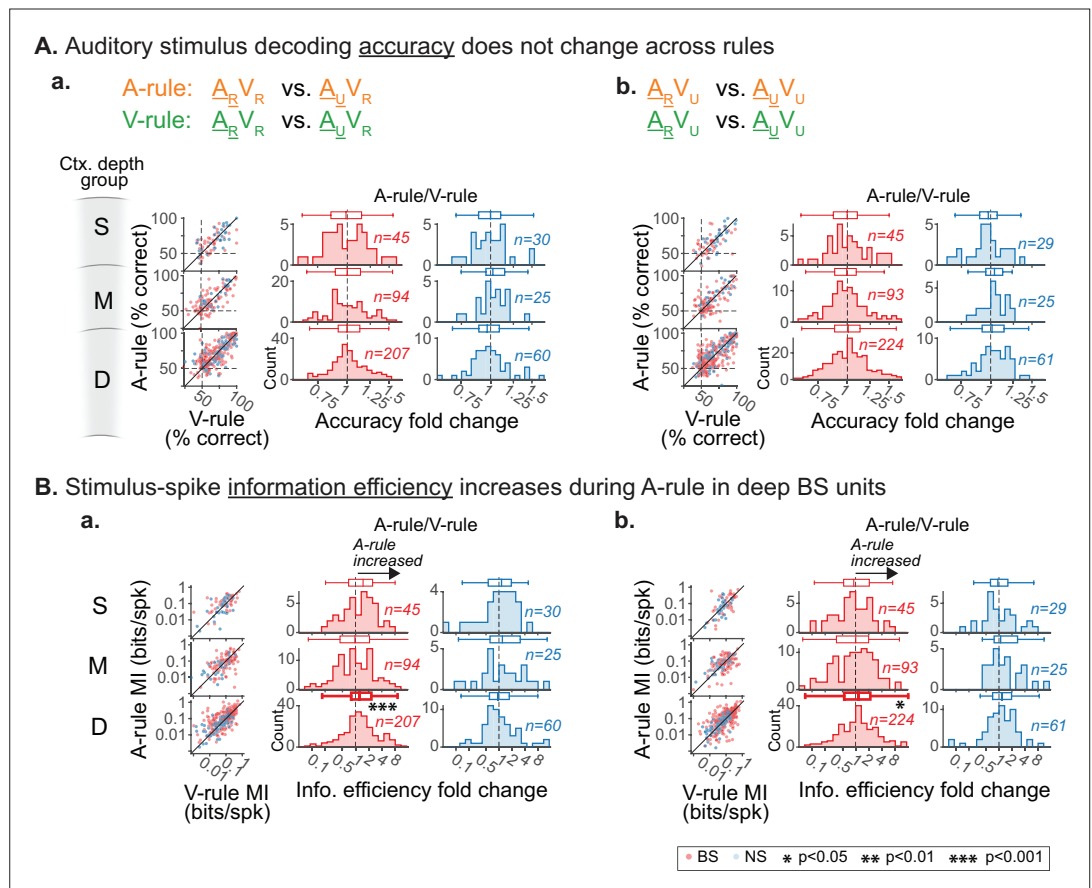


## Figure 7 continued

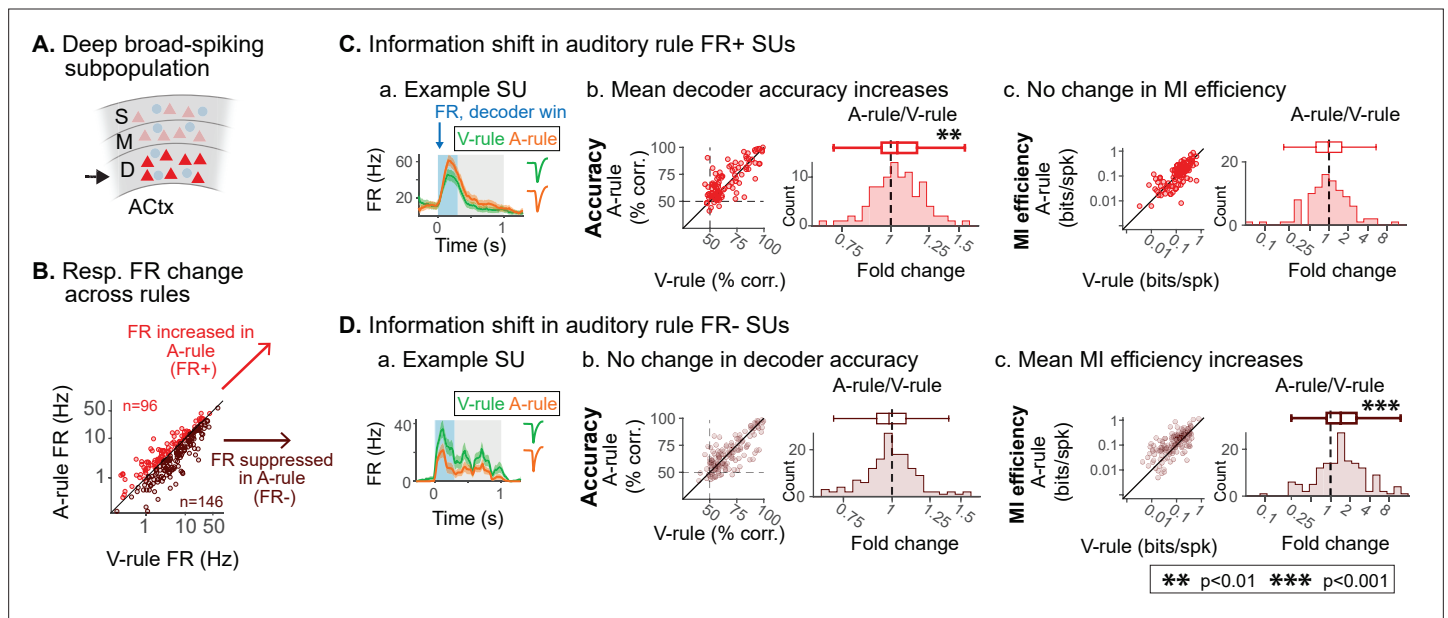
mutual information (MI; bits), and encoding efficiency (bits/spike). **(B)** Decoding accuracy of auditory stimulus identity, compared across attentional states. Decoder setup mimics task faced by mice in the A-rule: discrimination between  $A_R^*$  and  $A_U^*$  tone cloud identities. Results represent average of two decoder runs, which differ in their paired visual stimulus, but yield similar results:  $A_R V_R$  vs.  $A_U V_R$  and  $A_R V_U$  vs.  $A_U V_U$  (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 spectrotemporal receptive fields (STRFs) calculated from ITIs of A-rule and V-rule from the same single unit (SU). **(b)** Estimation of mutual information efficiency of ITI random double sweep (RDS) stimuli: the STRF is convolved with the windows of the RDS stimulus to define two distributions of relative STRF-stimulus similarity values: 1.  $P(x|\text{spike})$ , 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. Mutual information (MI) values are calculated from STRFs in A-rule and V-rule separately. **(E)** Example of spike train-STRF encoding efficiency from two SUs: low **(a)** and high **(b)** bits/spike examples. **(F)** Comparison of spike train-STRF encoding efficiency across rules, showing increased encoding efficiency in A-rule for deep broad-spiking units.

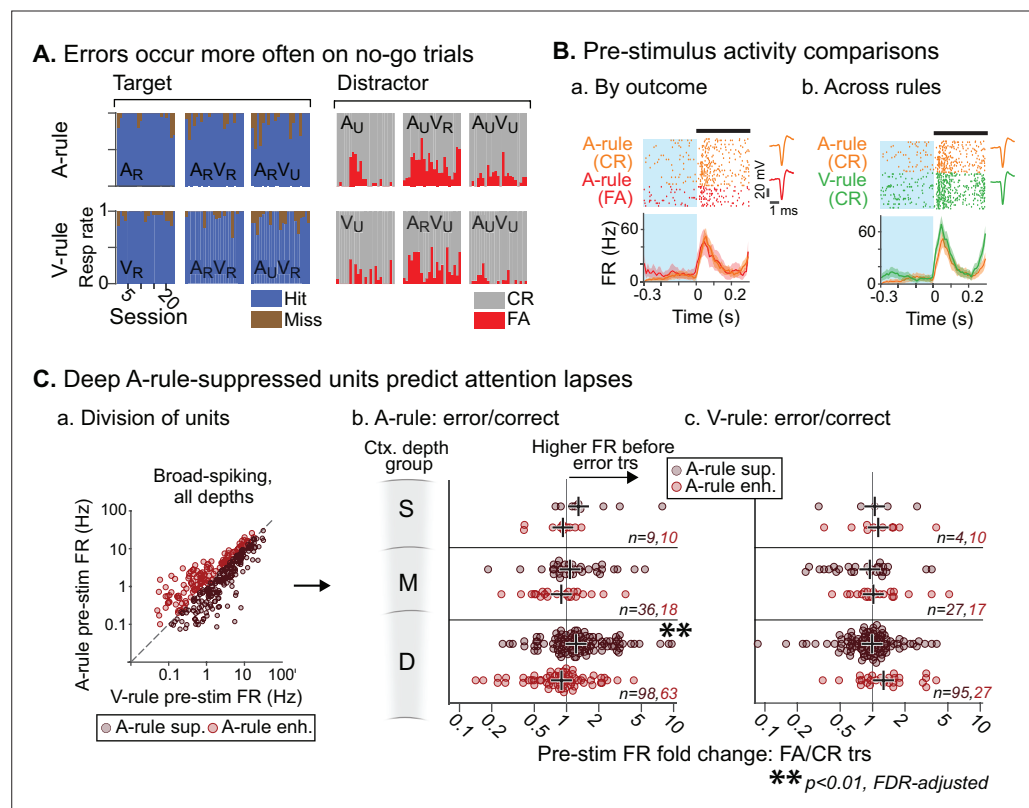


**Figure 7—figure supplement 1.** Decoding of task rule from stimulus response peristimulus time histograms (PSTHs). **(A)** Decoding of rule identity from PSTH responses using Euclidean distance-based classifier. Responses to each of the four audiovisual (AV) stimuli in the A-rule vs. responses to the 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 single unit (SU) decoding accuracy (% correct); broad-spiking (BS) units in red, narrow-spiking (NS) in blue. Chance decoding is indicated by the dashed line. A minimum 1 Hz firing rate (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: BS units, right: NS. 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)**  $A_R V_R$  vs.  $A_U V_R$ : no difference across rules (all  $p \geq 0.36$ , all  $|Z| \leq 0.92$ , paired Wilcoxon signed rank (WSR) test, see **Figure 7—source data 3**). **(b)**  $A_R V_U$  vs.  $A_U V_U$ : no difference across rules (all  $p \geq 0.24$ , all  $|Z| \leq 1.17$ ). 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)**  $A_R V_R$  vs.  $A_U V_R$ : deep group shows increased information efficiency in A-rule (deep BS:  $p = 5.8 \times 10^{-4}$ ,  $Z = 1.2$ , paired Wilcoxon signed-rank (WSR) test, p-values FDR-corrected; see **Figure 7—source data 4** for other groups) **(b)**  $A_R V_U$  vs.  $A_U V_U$ : deep BS group also shows increase in A-rule ( $p = 0.036$ ,  $Z = 1.15$ ).





**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 FAs vs. correct rejects (CRs). **(B)** Example unit showing behavioral outcome- and rule-dependent firing rate (FR) modulation. Pre-stimulus FR analysis window ( $-0.3$ – $0$  s) shown in blue. **(a)** Pre-stimulus activity for A-rule FA trials (red) is elevated relative to CR trials (orange). **(b)** In the same unit, pre-stimulus 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)** Broad-spiking (BS) units from sessions with  $\geq 10$  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, false discovery rate [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).