Neuroscience

Auditory Cortex Learns to Discriminate Audiovisual Cues through Selective Multisensory Enhancement

Reviewed Preprint
v1 • November 20, 2024
Not revised

Song Chang, Beilin Zheng, Les Keniston, Jinghong Xu, Liping Yu

Key Laboratory of Brain Functional Genomics (Ministry of Education and Shanghai), School of Life Sciences, East China Normal University, Shanghai, China • College of Information Engineering, Hangzhou Vocational & Technical College, Hangzhou, China • Department of Biomedical Sciences, Kentucky College of Osteopathic Medicine, University of Pikeville, Pikeville, United States

eLife Assessment

This is an **important** study that aims to investigate the behavioral relevance of multisensory responses recorded in the auditory cortex. The experiments are elegant and well-designed, and are supported by appropriate analyses of the data. However, the evidence presented for learning-dependent encoding of visual information is **incomplete** and it is possible that the surprisingly short-latency increases in activity are actually motor-related signals. Demonstrating that they really are visual responses is necessary in order to draw definitive conclusions from this study.

https://doi.org/10.7554/eLife.102926.1.sa3

Abstract

Multisensory object discrimination is essential in everyday life, yet the neural mechanisms underlying this process remain unclear. In this study, we trained rats to perform a two-alternative forced-choice task using both auditory and visual cues. Our findings reveal that multisensory perceptual learning actively engages auditory cortex (AC) neurons in both visual and audiovisual processing. Importantly, many audiovisual neurons in the AC exhibited experience-dependent associations between their visual and auditory preferences, displaying a unique integration model. This model employed selective multisensory enhancement for specific auditory-visual pairings, which facilitated improved multisensory discrimination. Additionally, AC neurons effectively distinguished whether a preferred auditory stimulus was paired with its associated visual stimulus using this distinct integrative mechanism. Our results highlight the capability of sensory cortices to develop sophisticated integrative strategies, adapting to task demands to enhance multisensory discrimination abilities.

Introduction

In our daily lives, the integration of visual and auditory information is crucial for detecting, discriminating, and identifying multisensory objects. When faced with stimuli like seeing an apple while hearing the sound 'apple' or 'peach' simultaneously, our brain must synthesize these inputs to form perceptions and make judgments based on the knowledge of whether the visual information matches the auditory input. To date, it remains unclear exactly where and how the brain integrates across-sensory inputs to benefit cognition and behavior. Traditionally, higher association areas in the temporal, frontal, and parietal lobes were thought to be pivotal for merging visual and auditory signals. However, recent research suggests that even primary sensory cortices, such as auditory and visual cortices, contribute significantly to this integration process[1-7].

Studies have shown that visual stimuli can modulate auditory responses in the AC via pathways involving the lateral posterior nucleus of the thalamus[8], and the deep layers of the AC serve as crucial hubs for integrating cross-modal contextual information[9]. Even irrelevant visual cues can affect how the AC perceives sound frequency[10]. Anatomical investigations reveal reciprocal nerve projections between auditory and visual cortices[4,11-14], highlighting the interconnected nature of these sensory systems. Sensory experience has been shown to shape cross-modal presentations in sensory cortices[15-17]. However, despite these findings, the precise mechanisms by which sensory cortices integrate cross-sensory inputs for multisensory object discrimination remain unknown.

Previous research on cross-modal modulation has predominantly focused on anesthetized or passive animal models[3,5,6,18,19], exploring the influence of stimulus properties and spatiotemporal arrangements on sensory interactions[3,6,19-21]. However, sensory representations, including multisensory processing, are known to be context-dependent[22-24], and can be shaped by perceptual learning [17,25]. Therefore, the way sensory cortices integrate information during active tasks may differ significantly from what has been observed in passive or anesthetized states. Relatively few studies have investigated cross-modal interactions during the performance of multisensory tasks, regardless of the brain region studied [25-27]. This limits our understanding of multisensory integration in sensory cortices, particularly regarding: (1) Do neurons in sensory cortices use the same or different strategies to integrate various audiovisual pairings? (2) How does the "match" or "mismatch" between auditory and visual features influence this integration? (3) How does learning to discriminate audiovisual objects affect their representations in sensory cortices?

We investigated this by training rats on a multisensory discrimination task involving both auditory and visual stimuli. We then examined cue selectivity and auditory-visual integration in AC neurons of these well-trained rats. Our findings demonstrate that multisensory discrimination training fosters experience-dependent associations between auditory and visual features within AC neurons. During task performance, AC neurons often exhibited multisensory enhancement for the preferred auditory-visual pairing, with no such enhancement observed for the non-preferred pairing. Importantly, this selective enhancement correlated with the animals' ability to discriminate the audiovisual pairings. Furthermore, the degree of auditory-visual integration correlated with the congruence of auditory and visual features. Our findings suggest AC plays a more significant role in multisensory integration than previously thought.

Results

Multisensory discrimination task in freely moving rats

To investigate how AC neurons integrate visual information into audiovisual processing, we trained ten adult male Long Evans rats on a multisensory discrimination task (**Fig. 1a**). During the task, the rat initiated a trial by inserting its nose into the central port, which triggered a randomly selected target stimulus from a pool of six cues: two auditory (3 kHz pure tone, A_{3k};10 kHz pure tone, A_{10k}), two visual (horizontal light bar, V_{hz}; vertical light bar, V_{vt}), and two multisensory cues ($A_{3k}V_{hz}$, $A_{10k}V_{vt}$). Based on the cue, the rats had to choose the correct left or right port for a water reward within 3 seconds after the stimulus onset. Incorrect choices or no response resulted in a 5-second timeout. To ensure reliable performance, rats needed to achieve 80% accuracy (including at least 70% correct in each modality) for three consecutive sessions (typically taking 2-4 months to achieve this level of accuracy).

Consistent with previous studies [10,26,28,29], rats performed better when responding to combined auditory and visual cues (multisensory trials) compared to trials with only auditory or visual cue (**Fig. 1b**). This suggests that multisensory cues facilitate decision-making. Reaction time, measured as the time from cue onset to when the rat left the central port, was shorter in multisensory trials (**Fig. 1c, d**). This indicates that multisensory processing in the brain helps rats discriminate between cues more efficiently (mean reaction time across rats, multisensory, 367±53 ms; auditory, 426±60 ms; visual, 417±53 ms; multisensory vs. each unisensory, $p < 0.001$, paired t-test. **Fig. 1d**).

Auditory, visual, and multisensory discrimination of AC neurons in multisensory discrimination task

To investigate the discriminative and integrative properties of AC neurons, we implanted tetrodes in the right primary auditory cortex of well-trained rats (n=10) (**Fig. 2a**). Careful implantation procedures were designed and followed to minimize neuron sampling biases. The characteristic frequencies of recorded AC neurons, measured immediately after tetrode implantation, spanned a broad frequency range (**Supplementary Fig. 1**). We examined a total of 559 AC neurons (56±29 neurons per rat) that responded to at least one target stimulus during task engagement. Interestingly, a substantial proportion of neurons (35%, 196/559) showed visual responses (**Fig. 2b**), which was notably higher than the 14% (14%, 39/275, $\chi 2 = 27.5$, $P < 0.001$) recorded in another group of rats (n=8) engaged in a choice-free task where rats were not required to discriminate triggered cues and could receive water rewards with any behavioral response (**Fig. 2b**). This suggests multisensory discrimination training enhances visual representation in the auditory cortex. Notably, 27% (150/559) of neurons responded to both auditory and visual stimuli (audiovisual neurons), and a small number (n=7) only responded to the auditory-visual combination (**Fig. 2b**).

During task engagement, AC neurons displayed a clear preference for one target sound over the other. As exemplified in **Fig. 2c**-**2d**, many neurons exhibited a robust response to one target sound while showing a weak or negligible response to the other. We quantified this preference using receiver operating characteristic (ROC) analysis. Since most neurons exhibited their main cue-evoked response within the initial period of cue presentation (**Fig. 2e**), our analysis focused on responses within the 0-150ms window after cue onset. We found a significant majority (61%, 307/506) of auditory-responsive neurons exhibited this selectivity during the task (**Fig. 2f**). Notably, most neurons favored the high-frequency tone (A_{10k} preferred: 261; A_{3k} preferred: 46, **Fig. 2f**). This aligns with findings that neurons in the AC and medial prefrontal cortex selectively preferred the tone associated with the behavioral choice contralateral to the recorded

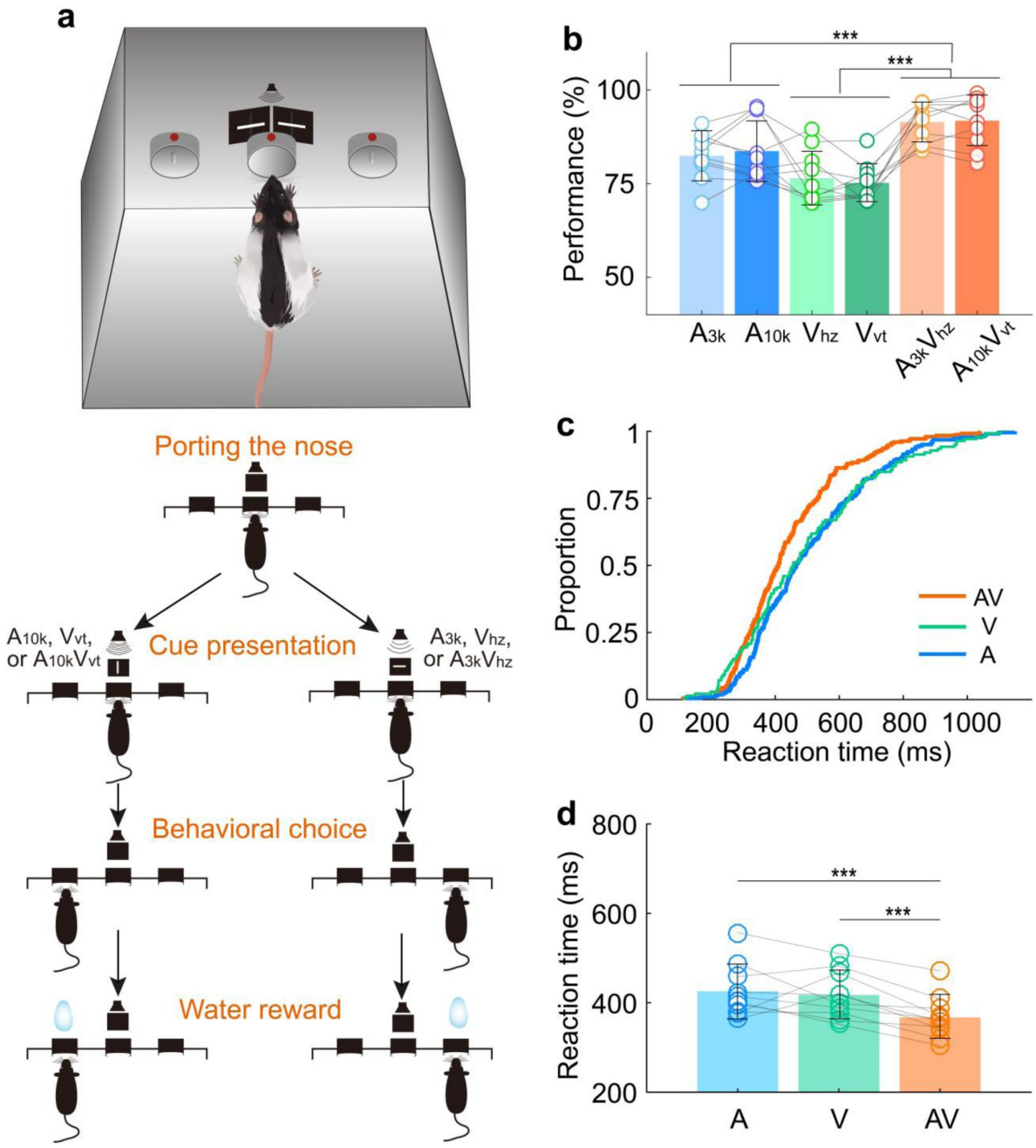


Fig. 1

Multisensory discrimination task for rats

a Schematic illustration of the behavioral task. A randomly selected stimulus was triggered when a rat placed its nose in the central port. If the triggered stimulus was a 10 kHz pure tone (A_{10k}), a vertical light bar (V_{vt}), or their combination ($A_{10k}V_{vt}$), the rat was rewarded at the left port. Conversely, for other stimuli (a 3 kHz pure tone (A_{3k}), a horizontal light bar (V_{hz}), or $A_{3k}V_{hz}$), the rat was required to move to the right port. Visual stimuli were presented via custom-made LED matrix panels, with one panel for each side. Auditory stimuli were delivered through a centrally located speaker. **b** The mean performance for each stimulus condition across rats. Circles connected by a gray line represent data from one individual rat. **c** Cumulative frequency distribution of reaction time (time from cue onset to leaving the central port) for one representative rat in auditory, visual and multisensory trials. **d** Comparison of average reaction times across rats in auditory, visual, and multisensory trials. ***, p<0.001. Error bars represent SDs.

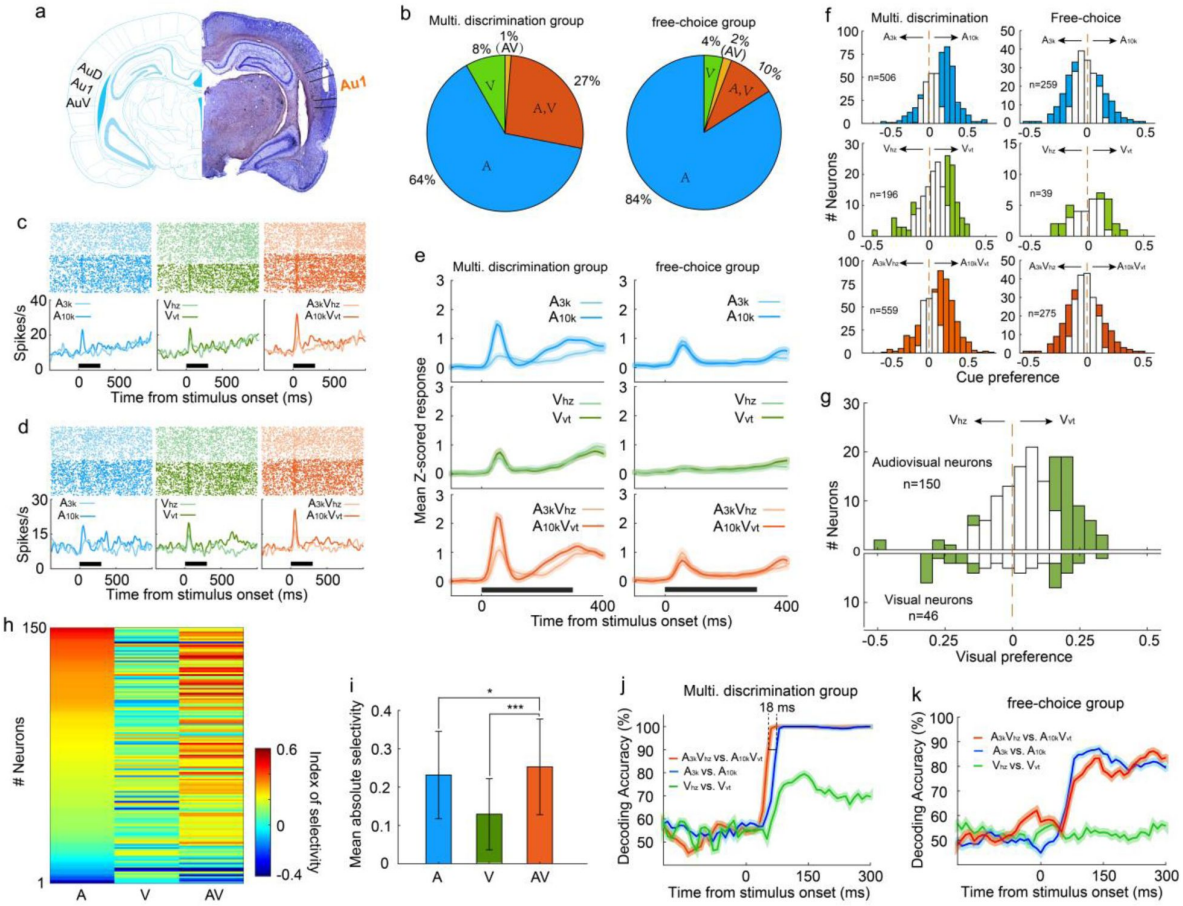


Fig. 2

Auditory, visual and multisensory selectivity in task engagement

a Histological verification of recording locations within the primary auditory cortex (Au1). **b** Proportion of neurons responding to auditory-only (A), visual-only (V), both auditory and visual (A, V), and audiovisual-only (VA) stimuli. **c-d** Rasters (top) and peristimulus time histograms (PSTHs, bottom) show responses of two exemplar neurons in A_{3k} (light blue), A_{10k} (dark blue), V_{hz} (light green), V_{vt} (dark green), $A_{3k}V_{hz}$ (light orange) and $A_{10k}V_{vt}$ (dark orange) trials (correct only). Mean spike counts in PSTHs were computed in 10-ms time windows and then smoothed with a Gaussian kernel (σ = 100ms). Black bars indicate the stimulus onset and duration. **e** Mean normalized response PSTHs across neurons in auditory (top), visual (middle), and multisensory (bottom) trials (correct only) for multisensory discrimination (left) and free-choice (right) groups. Shaded areas represent mean ± SEM. Black bars indicate stimulus onset and duration. **f** Histograms of auditory (top), visual (middle) and multisensory (bottom) selectivity for multisensory discrimination (left) and free-choice (right) groups. Filled bars indicate neurons for which the selectivity index was significantly different from 0 (permutation test, $p<0.05$, bootstrap n = 5000). The dash line represents zero. **g** Comparison of visual selectivity distribution between audiovisual (top) and visual (bottom) neurons. **h** Comparison of auditory (A), visual (V), and multisensory (AV) selectivity of 150 audiovisual neurons, ordered by auditory selectivity. **i** Average absolute auditory, visual and multisensory selectivity across 150 audiovisual neurons. Error bars represent SDs. *, $p<0.05$; ***, $p<0.001$. **j** Decoding accuracy of populations in the multisensory discrimination group. Decoding accuracy (cross-validation accuracy based on SVM) of populations between responses in two auditory (blue), two visual (green) and two multisensory (red) trials. Each decoding value was calculated in a 100ms temporal window moving at the step of 10ms. Shadowing represents the mean ± SD from bootstrapping of 100 repeats. Two dashed lines represent 90% of decoding accuracy for auditory and multisensory conditions. **k** Decoding accuracy of populations in the free-choice group.

cortices during sound discrimination tasks [10,30], potentially reflecting the formation of sound-to-action associations[31]. Such pronounced sound preference and bias were absent in the choice-free group (**Fig. 2f**-**2e**), suggesting it is directly linked to active discrimination. Anesthesia decreased auditory preference (**supplementary Fig. 2a**-**2b**), further supporting its dependence on active engagement.

Regarding the visual modality, 41% (80/196) of visually-responsive neurons showed a significant visual preference (**Fig. 2f**). Similar to the auditory selectivity observed, a greater proportion of neurons favored the visual stimulus (V_{vt}) associated with the contralateral choice, with a ratio of 60:20 for V_{vt} preferred vs. V_{hz} preferred neurons. This convergence of auditory and visual selectivity likely results from multisensory perceptual learning. Notably, such patterns were absent in neurons recorded from a separate group of rats performing choice-free tasks (**Fig. 2e**). Further supporting this conclusion is the difference in visual preference between audiovisual and exclusively visual neurons (**Fig. 2g**). Among audiovisual neurons with a significant visual selectivity, the majority favored V_{vt} (V_{vt} preferred vs. V_{hz} preferred, 48 vs. 7) (**Fig. 2g**), aligning with their established auditory selectivity. In contrast, visual neurons did not exhibit this bias (12 preferred V_{vt} vs. 13 preferred V_{hz}) (**Fig. 2g**). We propose that the dominant auditory input acts as a "teaching signal" that shapes visual processing through the selective reinforcement of specific visual pathways during associative learning. This aligns with Hebbian plasticity, where stronger auditory responses boost the corresponding visual input, ultimately leading visual selectivity to mirror auditory preference.

Similar to auditory selectivity, the vast majority of neurons (79%, 270/340) showing significant multisensory selectivity exhibited a preference for the multisensory cue ($A_{10k}V_{vt}$) guiding the contralateral choice (**Fig. 2f**). To more clearly highlight the influence of visual input on auditory selectivity, we compared auditory, visual, and multisensory selectivity in 150 audiovisual neurons (**Fig. 2h**). We found that pairing auditory cues with visual cues significantly improved the neurons' ability to distinguish between auditory stimuli alone (mean absolute auditory selectivity: 0.23±0.11; mean absolute multisensory selectivity: 0.25±0.12; $p<0.0001$, Wilcoxon Signed Rank Test; **Fig. 2i**).

Our multichannel recordings allowed us to decode sensory information from population activity in AC on a single-trial basis. Using cross-validated support vector machine (SVM) classifiers, we discriminated between auditory, visual, and multisensory cues. While decoding accuracy was similar for auditory and multisensory conditions, the presence of visual cues accelerated the decoding process, with AC neurons reaching 90% accuracy approximately 18ms earlier in multisensory trials (**Fig. 2j**), aligning with behavioral data. Interestingly, AC neurons could discriminate between two visual targets with around 80% accuracy (**Fig. 2j**), indicating a deeply accurate incorporation of visual processing in the auditory cortex. However, AC neurons in the free-choice group lacked visual discrimination ability and showed lower accuracy for auditory cues (**Fig. 2k**), suggesting that actively engaging multiple senses is crucial for the benefits of multisensory integration.

Audiovisual integration of AC neurons during the multisensory discrimination task

To understand how AC neurons integrate auditory and visual inputs during task engagement, we compared the multisensory response of each neuron to its strongest corresponding unisensory response using ROC analysis to quantify the difference, termed "modality selectivity". A selectivity greater than 0 indicates a stronger multisensory response. Over a third (34%, 192 of 559) of AC neurons displayed significant visual modulation of their auditory responses in one or both audiovisual pairings ($p < 0.05$, permutation test), including some neurons with no detectable visual response (104/356). **Fig. 3a** exemplifies this, where the multisensory response exceeded the auditory response, a phenomenon termed "multisensory enhancement" [32].

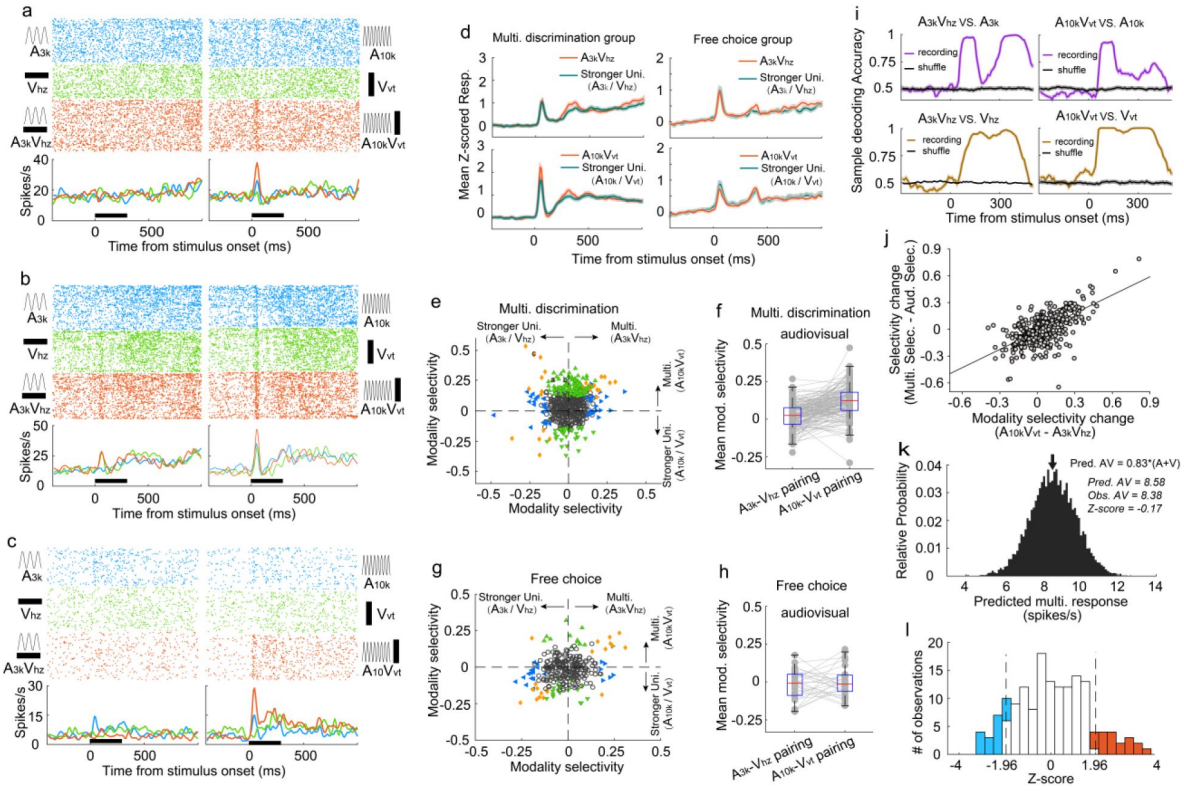


Fig. 3

Auditory and visual integration in multisensory discrimination task

a-c Rasters and PSTHs showing responses of three typical neurons recorded in well-trained rats performing multisensory discrimination tasks. **d** Population-averaged multisensory responses (red) compared to the corresponding stronger unisensory responses (dark blue) in A_{3k}-V_{hz} (top) and A_{10k}-V_{vt} (bottom) pairings for multisensory discrimination (left) and free-choice (right) groups. **e** Comparison of modality selectivity (multisensory vs. stronger unisensory) between A_{3k}-V_{hz} (x-axis) and A_{10k}-V_{vt} (y-axis) conditions. Each point represents values for a single neuron. Open circles: modality selectivity in either condition was not significant ($p > 0.05$, permutation test); triangles: significant selectivity in either A_{3k}-V_{hz} (blue) or A_{10k}-V_{vt} (green) condition; Diamonds: significant selectivity in both conditions. Dashed lines show zero modality selectivity. Points labeled a-c correspond to the neurons in panels a-c. **f** Mean modality selectivity for A_{10k}-V_{vt} and A_{3k}-V_{hz} pairings across audiovisual neurons in the multisensory discrimination group. **g** Comparison of modality selectivity for the free-choice group. **h** Mean modality selectivity across audiovisual neurons for the free-choice group. **i** SVM decoding accuracy in AC neurons between responses in multisensory vs. corresponding unisensory trials. Black line indicates shuffled control. **j** Positive relationship between the change in modality selectivity (A_{10k}-V_{vt} - A_{3k}-V_{hz}) and the selectivity change (multisensory selectivity - auditory selectivity). **k** Probability density functions of predicted mean multisensory responses (predicted AV) based on 0.83 times the sum of auditory (A) and visual (V) responses (same neuron as in **Fig. 2c**). The observed multisensory response matches the predicted mean (Z-score = -0.17). **l** Frequency distributions of *Z*-scores. Open bars indicate no significant difference between actual and predicted multisensory responses. Red bars: Z-scores>=1.96; blue bars: Z-scores<=-1.96.

Interestingly, AC neurons adopted distinct integration strategies depending on the specific auditory-visual pairing presented. Neurons often displayed multisensory enhancement for one pairing but not another (**Fig. 3b**), or even exhibited multisensory inhibition (**Fig. 3c**). This enhancement was mainly specific to the A_{10k}-V_{vt} pairing, as shown by population averages (**Fig. 3d**). Within this pairing, significantly more neurons exhibited enhancement than inhibition (114 vs. 36) (**Fig. 3e**). In contrast, the A_{3k}-V_{hz} pairing showed a balanced distribution of enhancement and inhibition (35 vs. 33) (**Fig. 3e**). This resulted in a significantly higher mean modality selectivity for the A_{10k}-V_{vt} pairing compared to the A_{3k}- V_{hz} pairing (0.047 ± 0.124 vs. 0.003 ± 0.096; paired t-test, $p < 0.001$). Among audiovisual neurons, this biasing is even more pronounced (enhanced vs. inhibited: 62 vs. 2 in A_{10k}-V_{vt} pairing, 6 vs. 13 in A_{3k}-V_{hz} pairing; mean modality selectivity: 0.119±0.105 in A_{10k}-V_{vt} pairing vs. 0.020±0.083 A_{3k}-V_{hz} pairing, paired t-test, p<-0.00001) (**Fig. 3f**). A similar but weaker pattern was observed under anesthesia (**Supplementary Fig. 2c**-**2f**), suggesting that multisensory perceptual learning may induce plastic changes in AC. In contrast, AC neurons in the choice-free group did not exhibit differential integration patterns (**Fig. 3g**-**3h**), indicating that multisensory discrimination and subsequent behavioral reporting are necessary for biased enhancement.

To understand how these distinct integration strategies influence multisensory discrimination, we compared the difference in modality selectivity between A_{10k}-V_{vt} and A_{3k}-V_{hz} pairings, and the change in neuronal multisensory versus auditory selectivity (**Fig. 3j**). We found a significant correlation between these measures (R=0.65, p<0.001, Pearson correlation test). The greater the difference in modality selectivity, the more pronounced the increase in cue discrimination during multisensory trials compared to auditory trials. This highlights how distinct integration models enhance cue selectivity in multisensory conditions. Additionally, a subset of neurons exhibited multisensory enhancement for each pairing (**Fig. 3e**), potentially aiding in differentiating between multisensory and unisensory responses. SVM decoding analysis confirmed that population neurons could effectively discriminate between multisensory and unisensory stimuli (**Fig. 3i**).

We further explored the integrative mechanisms—whether subadditive, additive, or superadditive — used by AC neurons to combine auditory and visual inputs. Using the bootstrap method, we generated a distribution of predicted multisensory responses by summing mean visual and auditory responses from randomly sampled trials where each stimulus was presented alone. Robust auditory and visual responses were primarily observed in correct contralateral choice trials, so our calculations focused on the A_{10k}-V_{vt} pairing. We found that, for most neurons, the observed multisensory response in contralateral choice trials was below the anticipated sum of the corresponding visual and auditory responses (**Supplementary Fig. 3**). Specifically, as exemplified in **Fig. 3k**, the observed multisensory response approximated 83% of the sum of the auditory and visual responses in most cases (**Fig. 3i**).

Impact of incorrect choices on audiovisual integration

To investigate how incorrect choices affected audiovisual integration, we compared multisensory integration for correct and incorrect choices within each auditory-visual pairing, focusing on neurons with a minimum of 9 trials per choice per cue. Our findings demonstrated a significant reduction in the magnitude of multisensory enhancement during incorrect choice trials in the A_{10k}-V_{vt} pairing. **Fig. 4a** illustrates a representative case. The mean modality selectivity for incorrect choices was significantly lower than that for correct choices (correct vs. wrong: 0.059±0.137 vs. 0.006±0.207; p=0.005, paired t-test, **Fig. 4 b-c**). This suggests that strong multisensory integration is crucial for accurate behavioral performance. In contrast, the A_{3k}-V_{hz} pairing showed no difference in modality selectivity between correct and incorrect trials (correct vs. wrong: 0.011±0.081 vs. 0.003±0.199; p=0.542, paired t-test, **Fig. 4d-e**). Interestingly, correct choices here likely correspond to ipsilateral behavioral selection, while incorrect choices correspond to contralateral behavioral selection. This indicates that contralateral behavioral

choice alone does not guarantee stronger multisensory enhancement. Overall, these findings suggest that the multisensory perception reflected by behavioral choices (correct vs. incorrect) might be shaped by the underlying integration strength.

Impact of informational match on audiovisual integration

A pivotal factor influencing multisensory integration is the precise alignment of informational content conveyed by distinct sensory cues. To explore the impact of the association status between auditory and visual cues on multisensory integration, we introduced two new multisensory cues ($A_{10k}V_{hz}$ and $A_{3k}V_{vt}$) into the target cue pool during task engagement. These cues were termed 'unmatched multisensory cues' as their auditory and visual components indicated different behavioral choices. Rats received water rewards with a 50% chance in either port when an unmatched multisensory cue was triggered.

We recorded from 280 AC neurons in well-trained rats performing a multisensory discrimination task with matched and unmatched audiovisual cues. We analyzed the integrative and discriminative characteristics of these neurons for matched and unmatched auditory-visual pairings. The results revealed distinct integrative patterns in AC neurons when an auditory target cue was paired with the matched visual cue as opposed to the unmatched visual cue. Neurons typically showed a robust response to A_{10k}, which was enhanced by the matched visual cue but was either unaffected (**Fig. 5a**) or inhibited (**Fig. 5b**) by the unmatched visual cue. In some neurons, both matched and unmatched visual cues enhanced the auditory response but matched cues provided a greater enhancement (**Fig. 5c**). We compared the modality selectivity for different auditory-visual pairings (**Fig. 5d**). Unlike V_{vt}, the unmatched visual cue, V_{hz}, generally failed to significantly enhance the response to the preferred sound (A_{10k}) in most cases (mean modality selectivity: 0.052 ± 0.097 for A_{10k}-V_{vt} pairing vs. 0.016 ± 0.123 for A_{10k}-V_{hz} pairing, $p<0.0001$, paired t-test). This suggests that consistent information across modalities strengthens multisensory integration. In contrast, neither associative nor non-associative visual cue could boost neurons' response to the nonpreferred sound (A_{3k}) overall (mean modality selectivity: 0.003 ± 0.008 for A_{3k}-V_{hz} pairing vs. 0.003± 0.006 for A_{3k}-V_{vt} pairing, p=0.19, paired t-test).

Although distinct AC neurons exhibited varying integration profiles for different auditory-visual pairings, we explored whether, as a population, they could distinguish these pairings. We trained a linear classifier to identify each pair of stimuli and employed the classifier to decode stimulus information from the population activity of grouped neurons. The resulting decoding accuracy matrix for discriminating pairs of stimuli was visualized (**Fig. 5e**). We found that the population of neurons could effectively discriminate two target multisensory cues. While matched and unmatched visual cues failed to differentially modulate the response to the nonpreferred sound (A_{3k}) at the single-neuron level, grouped neurons still could discriminate them with a decoding accuracy of 0.79 (**Fig. 5e**), close to the accuracy of 0.81 for discriminating between A_{10k}-V_{vt} and A_{10k}-V_{hz} pairings. This suggests that associative learning experiences enabled AC neurons to develop multisensory discrimination abilities. However, the accuracy for discriminating matched vs. unmatched cues was lower compared to other pairings (**Fig. 5e**).

Cue preference and multisensory integration in left AC

Our data showed that most neurons in the right AC preferred the cue directing the contralateral choice, regardless of whether it was auditory or visual. However, this could simply be because these neurons were naturally more responsive to those specific cues, not necessarily because they learned an association between the cues and the choice. To address this, we trained another 4 rats to perform the same discrimination task but recorded neuronal activity in left AC (**Fig. 6a**). We analyzed 193 neurons, of which about a third (31%, 60/193) responded to both auditory and visual cues (**Fig. 6b**). Similar to the right AC, the average response across neurons in the left AC preferred cues guiding the contralateral choice (**Fig. 6c**). However, in this case, the preferred

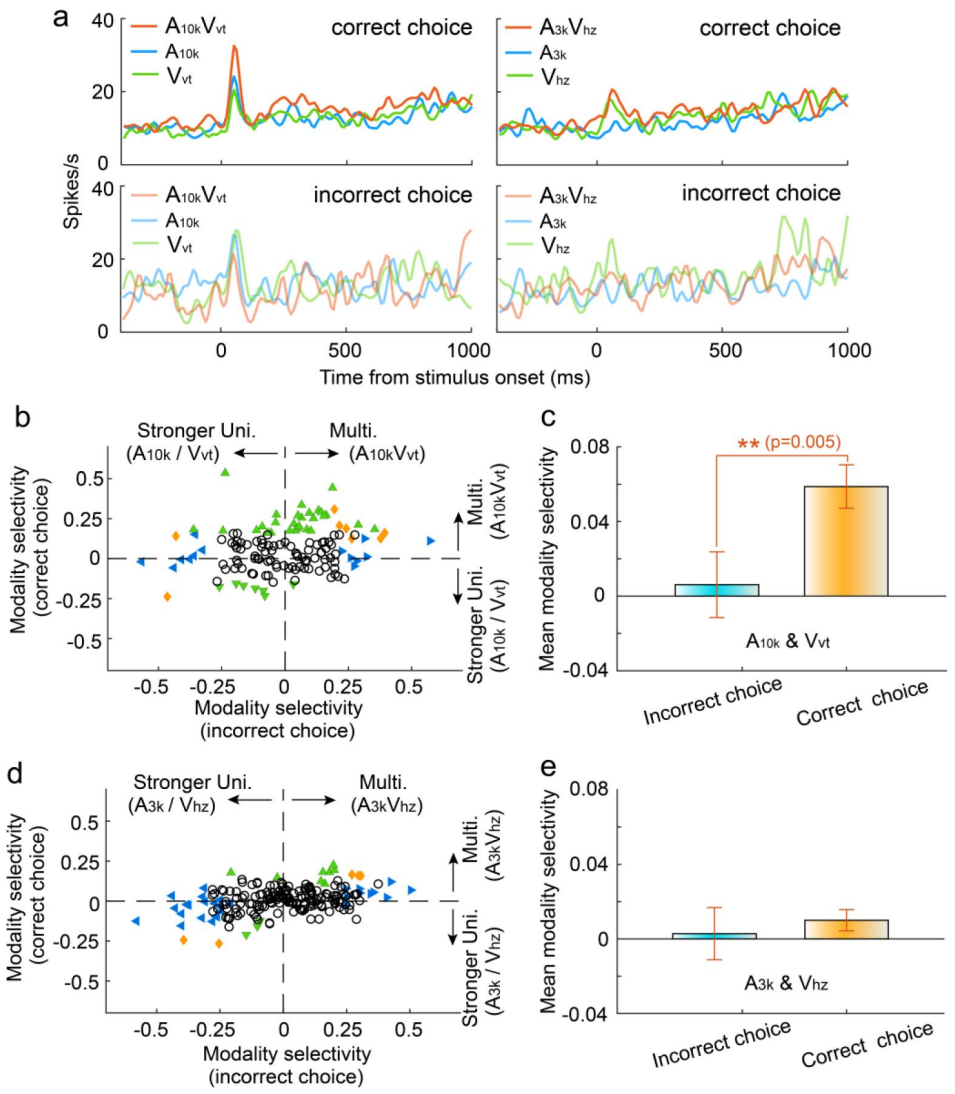


Fig. 4

Impact of Choice selection on Audiovisual Integration

a PSTHs show mean responses of an exemplar neuron for different cue and choice trials. **b** Mean modality selectivity across neurons for correct (orange) and incorrect (blue) choices in the A_{10k}-V_{vt} pairing. **c** Comparison of modality selectivity between correct and incorrect choices for the A_{10k}-V_{vt} pairing. Error bars, SEM. **d, e** Similar comparisons of modality selectivity for correct and incorrect choices in the A_{3k}-V_{hz} pairing.

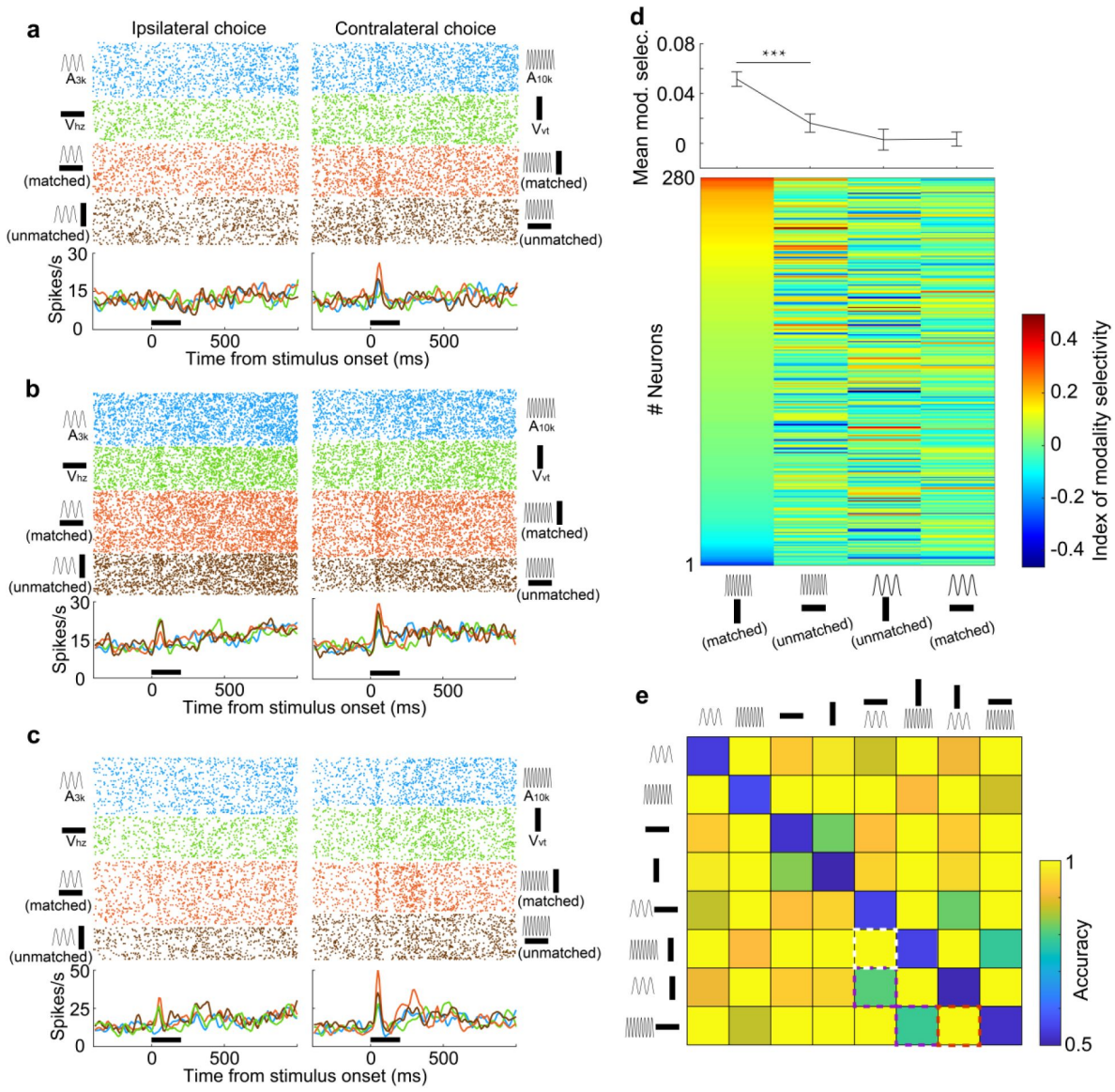


Fig. 5

Information-dependent integration and discrimination of audiovisual cues

a-c Responses of three example neurons to six target stimuli and two unmatched multisensory cues ($A_{3k}V_{vt}$ and $A_{10k}V_{hz}$) presented during task engagement. **d** Lower panel: Modality selectivity for different auditory-visual pairings. Neurons are sorted by their modality selectivity for the $A_{10k}V_{vt}$ condition. Upper panel: Mean modality selectivity across the population for each pairing. Error bar, SEM. ***, p<0.001. **e** Population decoding accuracy for all stimulus pairings (8×8) within a 150ms window after stimulus onset. White, purple, and red dashed squares denote the decoding accuracy for discriminating two target multisensory cues (white), discriminating two unmatched auditory-visual pairings (red) and matched versus unmatched audiovisual pairings (purple), respectively. Conventions are consistent with **Fig. 2**.

cues are A_{3k}, V_{hz} and the audiovisual pairing $A_{3k}V_{hz}$. As shown in **Fig. 6d**, more auditory-responsive neurons favored the sound denoting the contralateral choice. The same was true for visual selectivity in visually responsive neurons (**Fig. 6e**). This strongly suggests that the preference wasn't simply due to a general bias towards a specific cue, but rather reflected the specific associations the animals learned during the training.

Consistent with the cue biasing, differential multisensory integration was observed, with multisensory enhancement biased toward the A_{3k}-V_{hz} pairing (**Fig. 6f**). This mirrors the finding in the right AC, where multisensory enhancement was biased toward the auditory-visual pairing guiding the contralateral choice. In audiovisual neurons, mean modality selectivity in the A_{3k}-V_{hz} pairing is substantially higher than in the A_{10k}-V_{vt} pairing (0.135±0.126 vs. -0.039±0.138; $p<0.0001$, paired t-test) (**Fig. 6g**). These findings suggest that AC neurons exhibit cue preference and biased multisensory enhancement based on the learned association between cues and behavioral choice. This mechanism could enable the brain to integrate cues of different modalities into a common behavioral response.

Unisensory training does not replicate multisensory training effects

Our data suggest that most AC audiovisual neurons exhibited a visual preference consistent with their auditory preference following consistent multisensory discrimination training. Additionally, these neurons developed selective multisensory enhancement for a specific audiovisual stimulus pairing. To investigate whether these properties stemmed solely from long-term multisensory discrimination training, we trained a new group of animals (n=3) first on auditory and then on visual discriminations (**Fig. 7a**). These animals did not receive task-related multisensory associations during the training period. We then examined the response properties of neurons recorded in the right AC when well-trained animals performed auditory, visual and audiovisual discrimination tasks.

Among recorded AC neurons, 28% (49/174) responded to visual target cues (**Fig. 7b**). Unlike the multisensory training group, most visually responsive neurons in the unisensory training group lacked a visual preference, regardless of whether they were visually-only responsive or audiovisual (**Fig. 7c**). Interestingly, similar to the multisensory training group, nearly half of the recorded neurons (47%, 75/159) demonstrated clear auditory discrimination, with most (80%, 60 out of 75) favoring the sound guiding the contralateral choice (**Fig. 7d**). Furthermore, the unisensory training group did not exhibit population-level multisensory enhancement (n=174, **Fig. 7f**). The mean modality selectivity for the A_{3k}- V_{Hz} and A_{10k}-V_{Hz} pairs showed no significant difference ($p = 0.327$, paired t-test), with values of - 0.02±0.12 and -0.01±0.13, respectively (**Fig. 7g**). Even among audiovisual neurons (n=37), multisensory integration did not differ significantly between the two pairings (A_{3k}-V_{Hz} vs A_{10k}-V_{Hz}: 0.002±0.126 vs 0.017±0.147; $p=0.63$; **Fig. 7g**). These findings suggest that the development of multisensory enhancement for specific audiovisual cues and the alignment of auditory and visual preferences likely depend on the association of auditory and visual stimuli with the corresponding behavioral choice during multisensory training. Unisensory training alone cannot replicate these effects.

Discussion

In this study, we investigated how AC neurons integrate auditory and visual inputs to discriminate multisensory objects and whether this integration reflects the congruence between auditory and visual cues. Our findings reveal that most AC neurons exhibited distinct integrative patterns for different auditory-visual pairings, with multisensory enhancement observed primarily for favored pairings. This indicates that AC neurons show significant selectivity in their integrative processing. Furthermore, AC neurons effectively discriminated between matched and mismatched auditory-

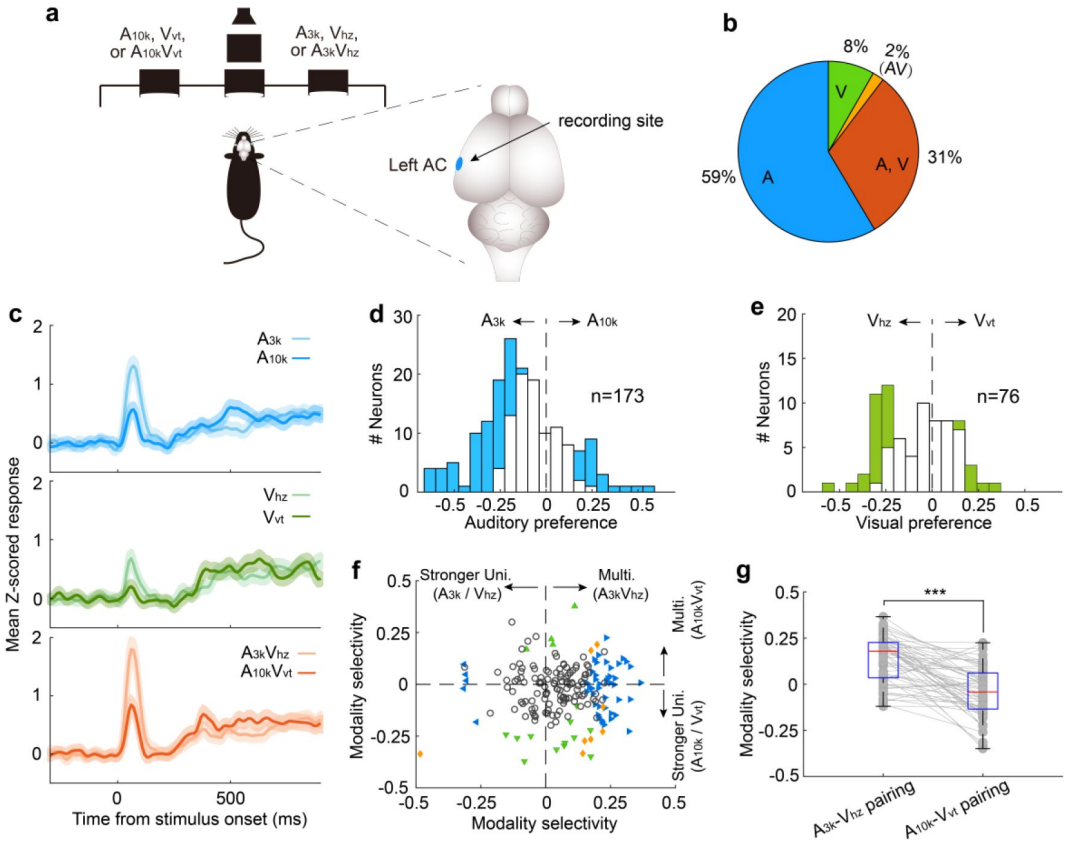


Fig. 6

Multisensory integration and cue preferences of neurons recorded in left AC

a The behavioral task and the recording site within the brain (the left AC). **b** The proportion of neurons responsive to auditory-only (A), visual-only (V), both auditory and visual (A, V), and audiovisual-only (VA) stimuli based on their spiking activity. **c** The mean PSTHs of normalized responses across neurons for different cue conditions. **d** A histogram of auditory selectivity index. Filled bars represent neurons with a significant selectivity index (permutation test, $P < 0.05$, bootstrap n = 5000). **e** Similar to panel **d**, this panel shows a histogram of the visual selectivity index. **f** Comparison of modality selectivity between A_{3k}-V_{hz} and A_{10k}-V_{vt} pairings. **g** A boxplot shows the comparison of modality selectivity for audiovisual neurons. ***, $P<0.001$. The conventions used are consistent with those in **Fig. 2** and **Fig. 3**.

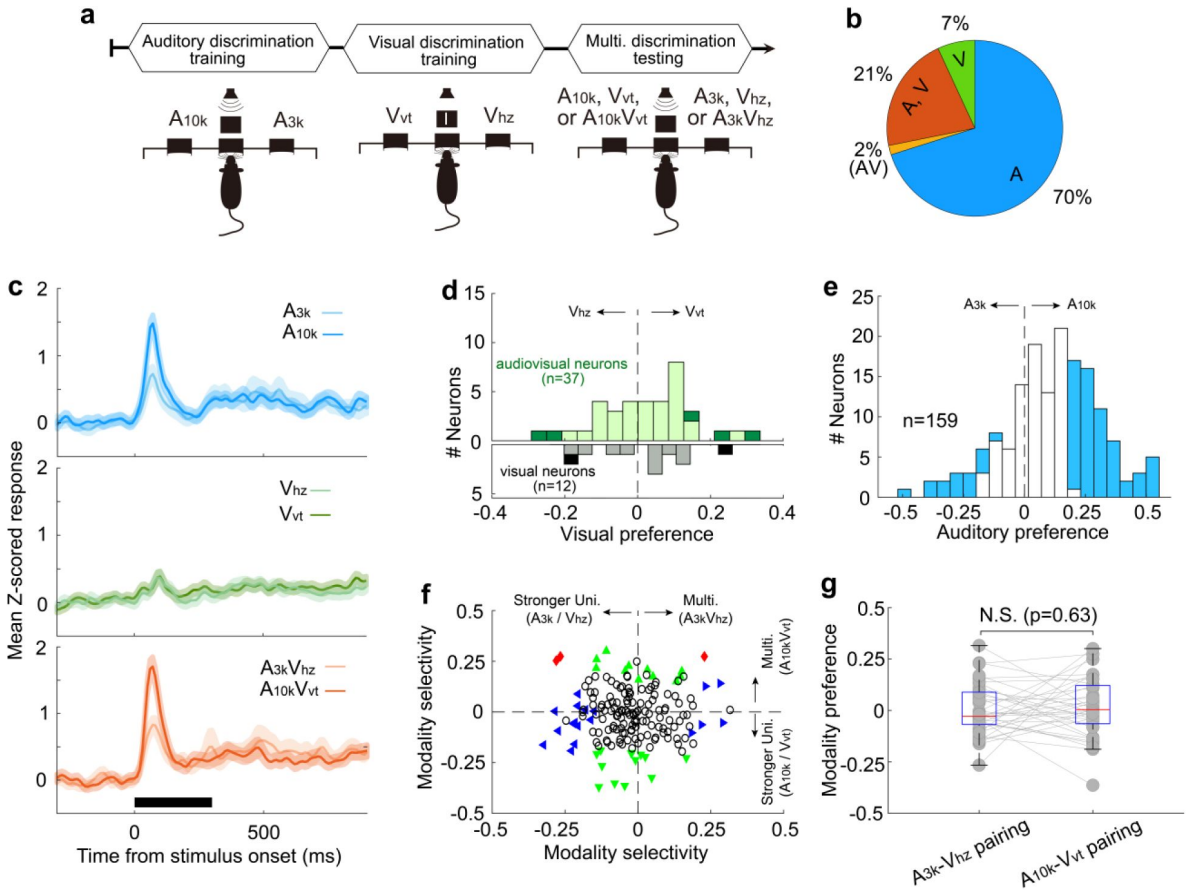


Fig. 7

Multisensory integration and cue preferences in right AC neurons after unisensory training

a This panel depicts the training stages for the rats: auditory discrimination training followed by visual discrimination training. **b** The proportion of neurons of different types. **c** Mean normalized responses across neurons for different cue conditions. **d-e** Histograms of visual (**d**) and auditory (**e**) selectivity. **f** Comparison of neuronal modality selectivity between A_{3k}-V_{hz} and A_{10k}-V_{vt} pairings. **g** Modality selectivity comparison in audiovisual neurons. N.S., no significant difference. The conventions used are consistent with those in **Fig. 2**, **Fig. 3** and **Fig. 6**.

visual pairings, highlighting the crucial role of the AC in multisensory object recognition and discrimination. Interestingly, a subset of auditory neurons not only developed visual responses but also exhibited congruence between auditory and visual selectivity. This suggests that multisensory perceptual training leads to auditory-visual binding and establishes a memory trace of the trained audiovisual experiences within the AC. Sensory cortices, Like AC, may act as a vital bridge for communicating sensory information across different modalities.

Numerous studies have explored the cross-modal interaction of sensory cortices[1,3,4,6,19,33]. Recent research has highlighted the critical role of cross-modal modulation in shaping the stimulus characteristics encoded by sensory cortical neurons. For instance, sound has been shown to refine orientation tuning in layer 2/3 neurons of the primary visual cortex[7], and a cohesive visual stimulus can enhance sound representation in the AC[5]. Despite these findings, the functional role and patterns of cross-modal interaction during perceptual discrimination remain unclear. In this study, we made a noteworthy discovery that AC neurons employed a nonuniform mechanism to integrate visual and auditory stimuli while well-trained rats performed a multisensory discrimination task. This differentially integrative pattern improved multisensory discrimination. Notably, this integrative pattern did not manifest during a free-choice task. These findings indicate that multisensory integration in sensory cortices is not static but rather task-dependent. We propose that task-related differentially integrative patterns may not be exclusive to sensory cortices but could represent a general model in the brain.

Consistent with prior research[10,30], the majority of AC neurons exhibited a selective preference for the cue linked to the contralateral choice, irrespective of the sensory modality. We propose that such cue-preference biasing helps animals learn associations between cues and actions. This learning may involve the formation of new connections between sensory and motor areas of the brain (cortico-cortical pathways) driven by sensorimotor association learning[34]. This biasing was not observed in the choice-free group. Similar biasing was also reported in a previous study, where auditory discrimination learning preferentially potentiated corticostriatal synapses from neurons representing either high or low frequencies associated with contralateral choices [31]. These findings underscore the significant impact of associative learning on cue encoding mechanisms, specifically highlighting the integration of stimulus discrimination and behavioral choice.

Prior work has investigated how neurons in sensory cortices discriminate unisensory cues in perceptual tasks[35-37]. It is well-established that when animals learn the association of sensory features with corresponding behavioral choices, cue representations in early sensory cortices undergo significant changes[24,38-40]. For instance, visual learning shapes cue-evoked neural responses and increases visual selectivity through changes in the interactions and correlations within visual cortical neurons [41,42]. Consistent with these previous studies, we found that auditory discrimination in the AC of well-trained rats substantially improved during the task.

In this study, we extended our investigation to include multisensory discrimination. We discovered that when rats performed the multisensory discrimination task, AC neurons exhibited robust multisensory selectivity, responding strongly to one auditory-visual pairing and showing weak or negligible responses to the other, similar to their behavior in auditory discrimination. Audiovisual neurons demonstrated higher selectivity in multisensory trials compared to auditory trials, driven by the differential integration pattern. Additionally, more AC neurons were involved in auditory-visual discrimination compared to auditory discrimination alone, suggesting that the recruitment of additional neurons may partially explain the higher behavioral performance observed in multisensory trials. A previous study indicates that cross-modal recruitment of more cortical neurons also enhances perceptual discrimination[43].

Our study explored how multisensory discrimination training influences visual processing in the AC. We observed well-trained rats exhibited a higher number of AC neurons responding to visual cues compared to untrained rats. This suggests that multisensory training enhances AC's ability to process visual information. Our result is in line with an increasing number of studies showing that multisensory perceptual learning can effectively drive plastic change in both sensory and association cortices[44,45]. For instance, a calcium imaging study in mice showed that a subset of multimodal neurons in mouse visual cortex develops enhanced auditory responses to the paired auditory stimulus following coincident auditory–visual experience[17]. A study conducted on the gustatory cortex of alert rats has shown that cross-modal associative learning was linked to a dramatic increase in the prevalence of neurons responding to nongustatory stimuli [16].

Among neurons responding to both auditory and visual stimuli, a congruent visual and auditory preference emerged during multisensory discrimination training, as opposed to unisensory discrimination training. These neurons primarily favored visual cues that matched their preferred sound. Interestingly, this preference was not observed in neurons solely responsive to visual targets. The strength of a visual response seems to be contingent upon the paired auditory input received by the same neuron. This aligns with known mechanisms in other brain regions, where learning strengthens or weakens connections based on experience[17,46]. This synchronized auditory-visual selectivity may be a way for AC to bind corresponding auditory and visual features, potentially forming memory traces for learned multisensory objects. These findings suggest that multisensory training may foster the formation of specialized neural circuits within AC. These circuits enable neurons to process related auditory and visual information together. However, further research is needed to determine if AC is the initial site for these circuit modifications.

There is ongoing debate about whether cross-sensory responses in sensory cortices predominantly reflect sensory inputs or are influenced by behavioral factors, such as cue-induced body movements. A recent study shows that sound-clip evoked activity in visual cortex have a behavioral rather than sensory origin and is related to stereotyped movements[47]. Several studies have demonstrated sensory neurons can encode signals associated with whisking[48], running[49], pupil dilation [50] and other movements[51]. In our study, however, we believe that the activity evoked by the flash of light bar in the AC reflects sensory inputs rather than behavioral modulation through visual-evoked body or orofacial movement. The observed responses to visual targets occurred mainly within a 100 ms temporal window following cue onset, which precedes the onset of rat body movement. This temporal dissociation strengthens the case for a sensory origin, aligning with other research showing motor-related activity triggered by sound begins much later in the visual cortex[52]. Additionally, movement evoked response are generally stereotyped, while the visual responses we observed are discriminative.

Additionally, our study sheds light on the role of semantic-like information in multisensory integration. During training, we created an association between specific auditory and visual stimuli, as both signaled the same behavioral choice. This setup mimics real-world scenarios where visual and auditory cues possess semantic coherence, such as an image of a cat paired with the sound of a "meow." Previous research has shown that semantically congruent multisensory stimuli enhance behavioral performance, while semantically incongruent stimuli either show no enhancement or result in decreased performance[53]. Intriguingly, our findings revealed a more nuanced role for semantic information. While AC neurons displayed multisensory enhancement for the preferred congruent audiovisual pairing, this wasn't observed for the other congruent pairing. This suggests that simple semantic similarity may not be enough for enhanced multisensory integration. However, the strength of multisensory enhancement itself served as a key indicator in differentiating between matched and mismatched cues. These findings provide compelling evidence that the nature of the semantic-like information plays a vital role in modifying multisensory integration at the neuronal level.

Methods

Animals

The animal procedures conducted in this study were ethically approved by the Local Ethical Review Committee of East China Normal University and followed the guidelines outlined in the Guide for the Care and Use of Laboratory Animals of East China Normal University. Twenty-five adult male Long-Evans rats, weighing approximately 250g, were obtained from the Shanghai Laboratory Animal Center (Shanghai, China) and used as subjects for the experiments. Of these rats, 14 were assigned to the multisensory discrimination task experiments, 3 to the unisensory discrimination task, and 8 to the control free-choice task experiments. The rats were group-housed with no more than four rats per cage and maintained on a regular light-dark cycle. They underwent water deprivation for two days before the start of behavioral training, with water provided exclusively inside the training box on training days. Training sessions were conducted six days per week, each lasting between 50 to 80 minutes, and were held at approximately the same time each day to minimize potential circadian variations in performance. The body weight of the rats was carefully monitored throughout the study, and supplementary water was provided to those unable to maintain a stable body weight from task-related water rewards.

Behavioral apparatus

All experiments were conducted in a custom-designed operant chamber, measuring 50×30×40 cm (length × width × height), with an open-top design. The chamber was placed in a sound-insulated double-walled room, and the inside walls and ceiling were covered with 3 inches of sound-absorbing foam to minimize external noise interference. One sidewall of the operant chamber was equipped with three snout ports, each monitored by a photoelectric switch (see **Fig. 1A**).

Automated training procedures were controlled using a real-time control program developed in MATLAB (Mathworks, Natick, MA, USA). The auditory signals generated by the program were sent to an analog-digital multifunction card (NI USB 6363, National Instruments, Austin, TX, USA), amplified by a power amplifier (ST-601, SAST, Zhejiang, China), and delivered through a speaker (FS Audio, Zhejiang, China). The auditory stimuli consisted of 300ms-long (15ms ramp-decay) pure tones with a frequency of 3kHz or 10kHz. The sound intensity was set at 60 dB sound pressure level (SPL) against an ambient background of 35-40 dB SPL. SPL measurements were taken at the position of the central port, which served as the starting position for the rats.

The visual cue was generated using two custom-made devices located on each side in front of the central port. Each device consisted of two arrays of closely aligned light-emitting diodes arranged in a cross pattern. The light emitted by the vertical or horizontal LED array passed through a ground glass, resulting in the formation of the corresponding vertical or horizontal light bar, each measuring 6×0.8 cm (**Fig. 1**). As stimuli, the light bars were illuminated for 300ms at an intensity of 2∼3cd/m^2. The audiovisual cue (multisensory cue) was the simultaneous presentation of both auditory and visual stimuli.

Multisensory discrimination task

The rats were trained to perform a cue-guided two-alternative forced-choice task, modified from previously published protocols [10,30]. Each trial began with the rat placing its nose into the center port. Following a short variable delay period (500-700ms) after nose entry, a randomly selected stimulus signal was presented. Upon cue presentation, the rats were allowed to initiate their behavioral choice by moving to either the left or right port (**Fig. 1A**). The training consisted of two stages. In the first stage, which typically lasted 3-5 weeks, the rats were trained to

discriminate between two audiovisual cues. In the second stage, an additional four unisensory cues were introduced, training the rats to discriminate a total of six cues (two auditory, two visual, and two audiovisual). This stage also lasted approximately 3-5 weeks.

During the task, the rats were rewarded with a drop (15-20μl) of water when they moved to the left reward port following the presentation of a 10 kHz pure tone sound, a vertical light bar, or their combination. For trials involving a 3 kHz tone sound, a horizontal light bar, or their combination, the rats were rewarded when they moved to the right reward port. Incorrect choices or failures to choose within 3 seconds after cue onset resulted in a timeout punishment of 5-6 seconds. Typically, the rats completed between 300 and 500 trials per day. They were trained to achieve a competency level of more than 80% correct overall and >70% correct in each cue condition in three consecutive sessions before the surgical implantation of recording electrodes.

The correct rate was calculated as follows:

$$\text{Correct rate (\%)} = 100*(\text{the number of correct trials}) / (\text{total number of trials});$$

The reaction time was defined as the temporal gap between the cue onset and the time when the rat withdrew its nose from the infrared beam in the cue port.

Unisensory discrimination task

Similar to the multisensory discrimination task, the rats first learned to discriminate between two auditory cues. Once their performance in the auditory discrimination task exceeded 75% correct, the rats then learned to discriminate between two visual cues. The auditory and visual cues used were the same as those in the multisensory discrimination task.

No cue discrimination choice-free task

In this task, the rats were not required to discriminate the cues presented. They received a water reward at either port after the onset of the cue. The six cues used were the same as those in the multisensory discrimination task. One week of training was sufficient for the rats to learn this no cue discrimination choice-free task.

Assembly of tetrodes

The tetrodes were constructed using Formvar-Insulated Nichrome Wire (bare diameter: 17.78 μm, A-M systems, WA, USA) twisted in groups of four. Two 20 cm-long wires were folded in half over a horizontal bar to facilitate twisting. The ends of the wires were clamped together and manually twisted in a clockwise direction. The insulation coating of the twisted wires was then fused using a heat gun at the desired twist level. Subsequently, the twisted wire was cut in the middle to produce two tetrodes. To enhance the longitudinal stability of each tetrode, it was inserted into polymide tubing (inner diameter: 0.045 inches; wall: 0.005 inches; A-M systems, WA, USA) and secured in place using cyanoacrylate glue. An array of 2×4 tetrodes was assembled, with an inter-tetrode gap of 0.4-0.5 mm. After assembly, the insulation coating at the tip of each wire was gently removed, and the exposed wire was soldered to a connector pin. For the reference electrode, a Ni-Chrome wire with a diameter of 50.8 μm (A-M systems, WA, USA) was used, with its tip exposed. A piece of copper wire with a diameter of 0.1 mm served as the ground electrode. Each of these electrodes was also soldered to a corresponding pin on the connector. The tetrodes, reference electrode, and ground electrode, along with their respective pins, were carefully arranged and secured using silicone gel. Immediately before implantation, the tetrodes were trimmed to an appropriate length.

Electrode Implantation

Prior to surgery, the animal received subcutaneous injections of atropine sulfate (0.01 mg/kg) to reduce bronchial secretions. The animal was then anesthetized with an intraperitoneal (i.p.) injection of sodium pentobarbital (40–50 mg/kg) and securely positioned on a stereotaxic apparatus (RWD, Shenzhen, China). An incision was made in the scalp, and the temporal muscle was carefully recessed. Subsequently, a craniotomy and durotomy were performed to expose the target brain region. Using a micromanipulator (RWD, Shenzhen, China), the tetrode array was slowly advanced and implanted in the right primary auditory cortex at coordinates 3.5 - 5.5 mm posterior to bregma and 6.4 mm lateral to the midline. The craniotomy was then sealed with tissue gel (3 M, Maplewood, MN, USA). The tetrode array was secured to the skull using stainless steel screws and dental acrylic. Following the surgery, animals received a 4-day prophylactic course of antibiotics (Baytril, 5 mg/kg, body weight, Bayer, Whippany, NJ, USA). They were allowed a recovery period of at least 7 days (typically 9-12 days) with free access to food and water.

Neural recordings

After rats had sufficiently recovered from the surgery, they resumed performing the same behavioral task they were trained to accomplish before surgery. Recording sessions were initiated once the animals' behavioral performance had returned to the level achieved before the surgery (typically within 2-3 days). Wideband neural signals in the range of 300-6000 Hz were recorded using the AlphaOmega system (AlphaOmega Instruments, Nazareth Illit, Israel). The amplified signals (×20) were digitized at a sampling rate of 25 kHz. These neural signals, along with trace signals representing the stimuli and session performance information, were transmitted to a PC for online observation and data storage.

Additionally, we recorded neural responses from well-trained rats under anesthesia. Anesthesia was induced with an intraperitoneal injection of sodium pentobarbital (40 mg/kg body weight) and maintained throughout the experiment by continuous intraperitoneal infusion of sodium pentobarbital (0.004 ~ 0.008 g/kg/h) using an automatic microinfusion pump (WZ-50C6, Smiths Medical, Norwell, MA, USA). The anesthetized rats were placed in the behavioral training chamber, and their heads were positioned in the cue port to mimic the cue-triggering as observed during task engagement. To maintain body temperature, a heating blanket was used to maintain a temperature of 37.5 °C. The same auditory, visual, and audiovisual stimuli used during task engagement were randomly presented to the anesthetized rats. For each cue condition, we recorded 40-60 trials of neural responses.

Analysis of electrophysiological data

The raw neural signals were recorded and saved for subsequent offline analysis. Spike sorting was performed using Spike 2 software (CED version 8, Cambridge, UK). Initially, the recorded raw neural signals were band-pass filtered in the range of 300-6000 Hz to eliminate field potentials. A threshold criterion, set at no less than three times the standard deviation (SD) above the background noise, was applied to identify spike peaks. The detected spike waveforms were then subjected to clustering using template-matching and principal component analysis. Waveforms with inter-spike intervals of less than 2.0 ms were excluded from further analysis. Spike trains corresponding to an individual unit were aligned to the onset of the stimulus and grouped based on different cue and choice conditions. Units were included for further analysis only if their overall mean firing rate within the session was at least 2 Hz. To generate peristimulus time histograms (PSTHs), the spike trains were binned at a resolution of 10 ms, and the average firing rate in each bin was calculated. The resulting firing rate profile was then smoothed using a Gaussian kernel with a standard deviation (σ) of 100 ms. In order to normalize the firing rate, the mean firing rate and SD during a baseline period (a 400-ms window preceding cue onset) were used to convert the averaged firing rate of each time bin into a Z-score.

Population decoding

To evaluate population discrimination for paired stimuli (cue_A vs. cue_B), we trained a Support Vector Machine (SVM) classifier with a linear kernel to predict cue selectivity. The SVM classifier was implemented using the "fitcsvm" function in Matlab. In this analysis, spike counts for each neuron in correct trials were grouped based on the triggered cues and binned into a 100 ms window with a 10 ms resolution. To minimize overfitting, only neurons with more than 30 trials for each cue were included. All these neurons were combined to form a pseudo population. The responses of the population neurons were organized into an M × N × T matrix, where M is the number of trials, N is the number of neurons, and T is the number of bins. For each iteration, 30 trials were randomly selected for each cue from each neuron. During cross-validation, 90% of the trials were randomly sampled as the training set, while the remaining 10% were used as the test set. The training set was used to compute the linear hyperplane that optimally separated the population response vectors corresponding to cue_A vs cue_B trials. The performance of the classifier was calculated as the fraction of correctly classified test trials, using 10-fold cross-validation procedures. To ensure robustness, we repeated the resampling process 100 times and computed the mean and standard deviation of the decoding accuracy across the 100 resampling iterations. Decoders were trained and tested independently for each bin. To assess the significance of decoding accuracy exceeding the chance level, shuffled decoding procedures were conducted by randomly shuffling the trial labels for 1000 iterations.

Cue selectivity

To quantify cue (auditory, visual and multisensory) selectivity between two different cue conditions (e.g., low tone trials vs. high tone trials), we employed a receiver operating characteristic (ROC) based analysis, following the method described in a previous study[54]. This approach allowed us to assess the difference between responses in cue_A and cue_B trials. Firstly, we established 12 threshold levels of neural activity that covered the range of firing rates obtained in both cue_A and cue_B trials. For each threshold criterion, we plotted the proportion of cue_A trials where the neural response exceeded the criterion against the proportion of cue_B trials exceeding the same criterion. This process generated an ROC curve, from which we calculated the area under the ROC curve (auROC). The cue selectivity value was then defined as 2 * (auROC - 0.5). A cue selectivity value of 0 indicated no difference in the distribution of neural responses between cue_A and cue_B, signifying similar responsiveness to both cues. Conversely, a value of 1 or -1 represented the highest selectivity, indicating that responses triggered by cue_A were consistently higher or lower than those evoked by cue_B, respectively. To determine the statistical significance of the observed cue selectivity, we conducted a two-tailed permutation test with 2000 permutations. We randomly reassigned the neural responses to cue_A and cue_B trials and recalculated a cue selectivity value for each permutation. This generated a distribution of values from which we calculated the probability of our observed result. If our actual value fell within the top 5% of this distribution, it was deemed significant (i.e., $p < 0.05$).

Comparison of actual and predicted multisensory responses

We conducted a comprehensive analysis to compare the observed multisensory responses with predicted values. The predicted multisensory response is calculated as either the sum of visual and auditory responses or as a coefficient multiplied by this sum. To achieve this, we first computed the mean observed multisensory response by averaging across audiovisual trials. We then created a benchmark distribution of predicted multisensory responses by iteratively calculating all possible predictions. In each iteration, we randomly selected (without replacement) the same number of trials as used in the actual experiment for both auditory and visual conditions. The responses from these selected auditory and visual trials were then averaged to obtain mean predicted auditory and visual responses, which were used to create a predicted multisensory response. This process was repeated 5,000 times to generate a comprehensive reference

distribution of predicted multisensory responses. By comparing the actual mean multisensory response to this distribution, we expressed their relationship as a Z-score. This method allowed us to quantitatively assess multisensory interactions, providing valuable insights into neural processing and enhancing our understanding of the mechanisms underlying multisensory integration.

Histology

Following the final data recording session, the precise tip position of the recording electrode was marked by creating a small DC lesion (-30 μA for 15 s). Subsequently, the rats were deeply anesthetized with sodium pentobarbital (100 mg/kg) and underwent transcardial perfusion with saline for several minutes, followed immediately by phosphate-buffered saline (PBS) containing 4% paraformaldehyde (PFA). The brains were carefully extracted and immersed in the 4% PFA solution overnight. To ensure optimal tissue preservation, the fixed brain tissue underwent cryoprotection in PBS with a 20% sucrose solution for at least three days. Afterward, the brain tissue was coronally sectioned using a freezing microtome (Leica, Wetzlar, Germany) with a slice thickness of 50 μm. The resulting sections, which contained the auditory cortex, were stained with methyl violet to verify the lesion sites and/or the trace of electrode insertion within the primary auditory cortex.

Statistical analysis

We also use ROC analysis to calculate the modality selectivity to denote the difference between multisensory and the corresponding stronger unisensory responses. All statistical analyses were conducted in MATLAB, and statistical significance was defined as a P value of < 0.05. To determine the responsiveness of AC neurons to sensory stimuli, neurons exhibiting evoked responses greater than 2 spikes/s within a 0.3 s window after stimulus onset, and significantly higher than the baseline response ($P < 0.05$, Wilcoxon signed-rank test), were included in the subsequent analysis. For behavioral data, such as mean reaction time differences between unisensory and multisensory trials, cue selectivity and mean modality selectivity across different auditory-visual conditions, comparisons were performed using either the Wilcoxon signed-rank test or paired t-test, as appropriate. We performed a Chi-square test to analyze the difference in the proportions of neurons responding to visual stimuli between the multisensory discrimination and free-choice groups. Correlation values were computed using Pearson's correlation. All data are presented as mean ± SD for the respective groups.

Acknowledgements

This work was supported by grants from the "STI2030-major projects" (2021ZD0202600), Natural Science Foundation of China (32371046, 31970925, 32271057).

Additional information

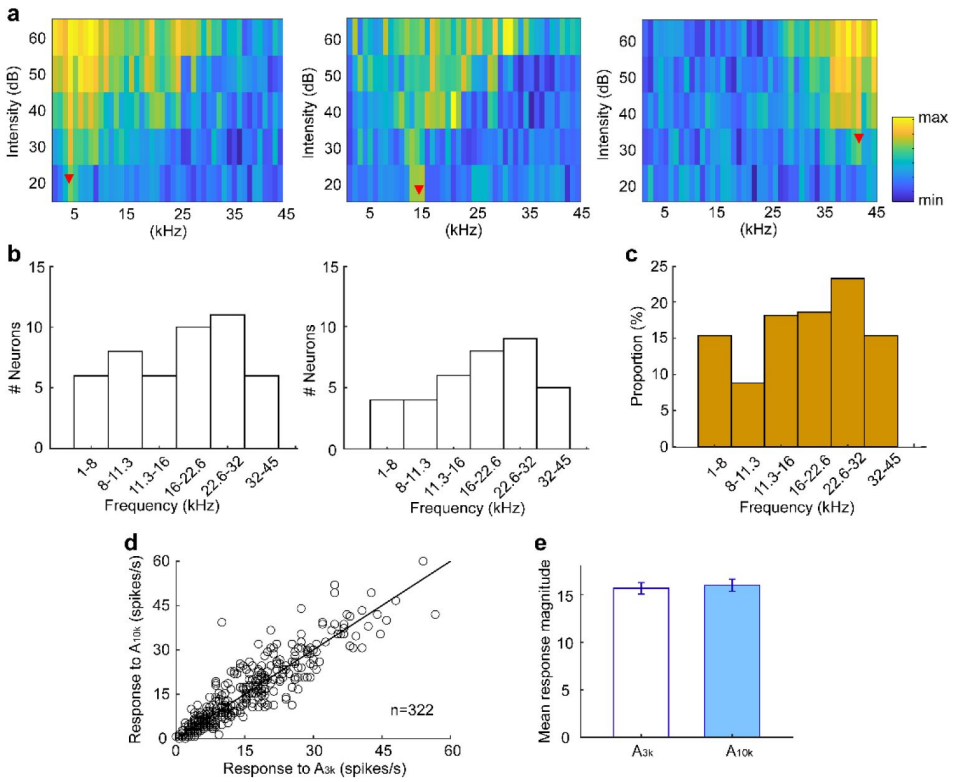
Author contributions

L.Y. and J.X. supervised and directed this project. L.Y. wrote the manuscript and revision. S.C. performed the experimental studies. J.X., L.K. and P.Z. helped with discussion of the study and contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

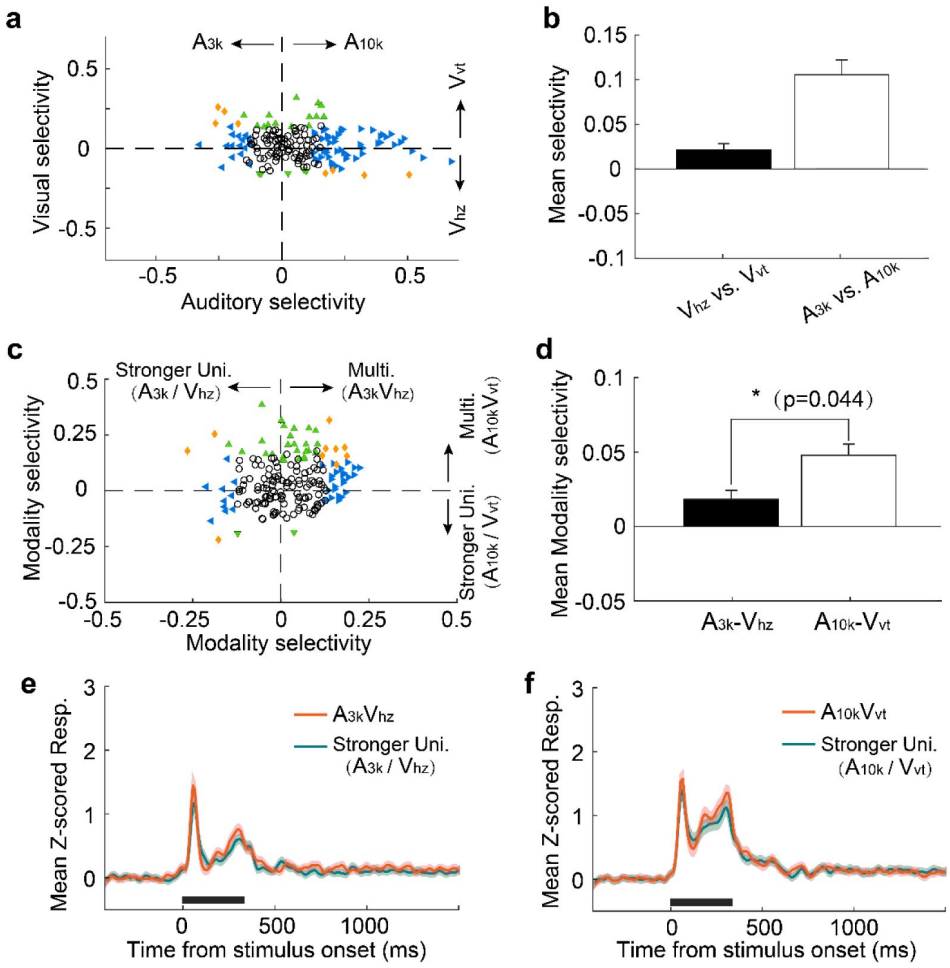
Supplemental figures



Supplementary Fig. 1

Characteristic Frequency (CF) and response to target sound stimuli of AC neurons recorded in well-trained rats under anesthesia

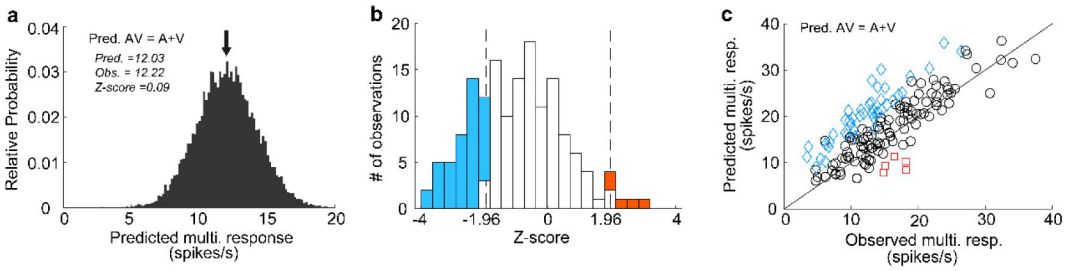
a Frequency-intensity tonal receptive fields (TRFs) of three representative AC neurons. The red triangle denotes CF. To create the TRF, we measured responses to a series of pure tone sounds (1-45 kHz in 1 kHz steps) at different intensities (20-60 dB in 10 dB increments). The color bar indicates the range of maximum and minimum spike response counts observed within the TRFs. **b** CF distributions of neurons recorded from two representative rats. **c** Overall distribution of all recorded neurons. **d** Comparison of responses to 3 kHz (A_{3k}) and 10 kHz (A_{10k}) pure tones. Each circle represents one neuron. **e** Comparison of mean response magnitude across populations for A_{3k} and A_{10k} conditions.



Supplementary Fig. 2

Cue and modality selectivity of AC neurons in well-trained rats under anesthesia

a Comparison of auditory and visual selectivity. Each point represents values for a single neuron. Open circles indicate neurons where neither auditory nor visual selectivity was significant ($p > 0.05$, permutation test). Triangles indicate significant selectivity in either auditory (blue) or visual (green) conditions, while diamonds represent significant selectivity in both auditory and visual conditions. Dashed lines represent zero cue selectivity. **b** Mean auditory (filled) and visual (unfilled) selectivity. Error bars represent SEMs. **c** Modality selectivity in A_{3k}-V_{hz} and A_{10k}-V_{vt} pairings. **d** Mean modality selectivity in A_{3k}-V_{hz} (unfilled) and A_{10k}-V_{vt} (filled) pairings. Error bars represent SEMs. **e, f** Comparison of mean multisensory responses (red) with mean corresponding largest unisensory responses (dark blue) across populations for A_{3k}-V_{hz} (**e**) and A_{10k}-V_{vt} (**f**) pairings. The conventions used are consistent with those in **Fig. 2** and **Fig. 3**.



Supplementary Fig. 3

The observed versus the predicted multisensory response

a Probability density functions of the predicted mean multisensory responses (predicted AV) based on summing the modality-specific visual (V) and auditory (A) responses (see *Methods* for details). Black arrow, the mean of the predicted distribution. The same neuron is shown in **Fig. 2B** and only responses in correct contralateral choice trials are involved. The predicted mean, actual mean observed, and their difference expressed in SDs (*Z*-score) are shown. **b** Frequency distributions of *Z*-scores. Open bars represent Z-scores between -1.96 and 1.96, indicating that the actual observed multisensory response is not significantly different from the predicted multisensory response (additive integration). Red bars, *Z*-score>=1.96 (superadditive integration); blue bars, *Z*-score<=-1.96 (subadditive integration). **c** Comparison between the mean predicted multisensory response and the actual observed multisensory response for all audiovisual neurons. Note that the predicted multisensory response is greater than the observed in most cases. Red squares and blue diamonds represent neurons with observed multisensory responses that are significantly greater (red squares) or less (blue diamonds) than the predicted mean multisensory responses.

References

1 Perrodin C., Kayser C., Logothetis N. K., Petkov C. I (2015) **Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex** *Proc Natl Acad Sci U S A* **112**:273–278 https://doi.org/10.1073/pnas.1412817112

2 Ghazanfar A. A., Maier J. X., Hoffman K. L., Logothetis N. K (2005) **Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex** *J Neurosci* **25**:5004–5012 https://doi.org/10.1523/JNEUROSCI.0799-05.2005

3 Kayser C., Petkov C. I., Logothetis N. K (2008) **Visual modulation of neurons in auditory cortex** *Cereb Cortex* **18**:1560–1574 https://doi.org/10.1093/cercor/bhm187

4 Bizley J. K., Nodal F. R., Bajo V. M., Nelken I., King A. J (2007) **Physiological and anatomical evidence for multisensory interactions in auditory cortex** *Cereb Cortex* **17**:2172–2189 https://doi.org/10.1093/cercor/bhl128

5 Atilgan H., et al. (2018) **Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding** *Neuron* **97**:640–655 https://doi.org/10.1016/j.neuron.2017.12.034

6 Meijer G. T., Montijn J. S., Pennartz C. M. A., Lansink C. S (2017) **Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency** *J Neurosci* **37**:8783–8796 https://doi.org/10.1523/JNEUROSCI.0468-17.2017

7 Ibrahim L. A., et al. (2016) **Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition** *Neuron* **89**:1031–1045 https://doi.org/10.1016/j.neuron.2016.01.027

8 Chou X. L., et al. (2020) **Contextual and cross-modality modulation of auditory cortical processing through pulvinar mediated suppression** *Elife* **9** https://doi.org/10.7554/eLife.54157

9 Morrill R. J., Bigelow J., DeKloe J., Hasenstaub A. R (2022) **Audiovisual task switching rapidly modulates sound encoding in mouse auditory cortex** *Elife* **11** https://doi.org/10.7554/eLife.75839

10 Chang S., et al. (2022) **Integrating Visual Information into the Auditory Cortex Promotes Sound Discrimination through Choice-Related Multisensory Integration** *J Neurosci* **42**:8556–8568 https://doi.org/10.1523/JNEUROSCI.0793-22.2022

11 Stehberg J., Dang P. T., Frostig R. D (2014) **Unimodal primary sensory cortices are directly connected by long-range horizontal projections in the rat sensory cortex** *Front Neuroanat* **8** https://doi.org/10.3389/fnana.2014.00093

12 Falchier A., Clavagnier S., Barone P., Kennedy H (2002) **Anatomical evidence of multimodal integration in primate striate cortex** *J Neurosci* **22**:5749–5759

13 Cappe C., Barone P (2005) **Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey** *Eur J Neurosci* **22**:2886–2902 https://doi.org/10.1111/j.1460-9568.2005.04462.x

14 Campi K. L., Bales K. L., Grunewald R., Krubitzer L (2010) **Connections of auditory and visual cortex in the prairie vole (Microtus ochrogaster): evidence for multisensory processing in primary sensory areas** *Cereb Cortex* **20**:89–108 https://doi.org/10.1093/cercor/bhp082

15 Xu J., Sun X., Zhou X., Zhang J., Yu L (2014) **The cortical distribution of multisensory neurons was modulated by multisensory experience** *Neuroscience* **272**:1–9 https://doi.org/10.1016/j.neuroscience.2014.04.068

16 Vincis R., Fontanini A (2016) **Associative learning changes cross-modal representations in the gustatory cortex** *Elife* **5** https://doi.org/10.7554/eLife.16420

17 Knopfel T., et al. (2019) **Audio-visual experience strengthens multisensory assemblies in adult mouse visual cortex** *Nat Commun* **10** https://doi.org/10.1038/s41467-019-13607-2

18 Stein B. E., Stanford T. R., Rowland B. A (2014) **Development of multisensory integration from the perspective of the individual neuron** *Nat Rev Neurosci* **15**:520–535 https://doi.org/10.1038/nrn3742

19 Deneux T., et al. (2019) **Context-dependent signaling of coincident auditory and visual events in primary visual cortex** *Elife* **8** https://doi.org/10.7554/eLife.44006

20 Wallace M. T., Woynaroski T. G., Stevenson R. A (2020) **Multisensory Integration as a Window into Orderly and Disrupted Cognition and Communication** *Annual Review of Psychology, Vol 71* **71**:193–219

21 Xu J., et al. (2018) **Spatial receptive field shift by preceding cross-modal stimulation in the cat superior colliculus** *J Physiol* **596**:5033–5050 https://doi.org/10.1113/JP275427

22 Elgueda D., et al. (2019) **State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex** *Nat Neurosci* **22**:447–459 https://doi.org/10.1038/s41593-018-0317-8

23 Han X., Xu J., Chang S., Keniston L., Yu L (2021) **Multisensory-Guided Associative Learning Enhances Multisensory Representation in Primary Auditory Cortex** *Cereb Cortex* https://doi.org/10.1093/cercor/bhab264

24 Otazu G. H., Tai L. H., Yang Y., Zador A. M (2009) **Engaging in an auditory task suppresses responses in auditory cortex** *Nat Neurosci* **12**:646–654 https://doi.org/10.1038/nn.2306

25 Garner A. R., Keller G. B (2022) **A cortical circuit for audio-visual predictions** *Nat Neurosci* **25**:98–105 https://doi.org/10.1038/s41593-021-00974-7

26 Raposo D., Kaufman M. T., Churchland A. K (2014) **A category-free neural population supports evolving demands during decision-making** *Nat Neurosci* **17**:1784–1792 https://doi.org/10.1038/nn.3865

27 Hirokawa J., et al. (2011) **Multisensory information facilitates reaction speed by enlarging activity difference between superior colliculus hemispheres in rats** *PLoS One* **6** https://doi.org/10.1371/journal.pone.0025283

28 Angelaki D. E., Gu Y., DeAngelis G. C (2009) **Multisensory integration: psychophysics, neurophysiology, and computation** *Curr Opin Neurobiol* **19**:452–458 https://doi.org/10.1016/j.conb.2009.06.008

29 Smyre S. A., Wang Z., Stein B. E., Rowland B. A (2021) **Multisensory enhancement of overt behavior requires multisensory experience** *Eur J Neurosci* **54**:4514–4527 https://doi.org/10.1111/ejn.15315

30 Zheng M., et al. (2021) **Choice-dependent cross-modal interaction in the medial prefrontal cortex of rats** *Mol Brain* **14** https://doi.org/10.1186/s13041-021-00732-7

31 Xiong Q., Znamenskiy P., Zador A. M (2015) **Selective corticostriatal plasticity during acquisition of an auditory discrimination task** *Nature* **521**:348–351 https://doi.org/10.1038/nature14225

32 Stein B. E., Meredith M. A. (1993) **The Merging of the Senses** MIT Press

33 Ghazanfar A. A., Schroeder C. E (2006) **Is neocortex essentially multisensory?** *Trends Cogn Sci* **10**:278–285 https://doi.org/10.1016/j.tics.2006.04.008

34 Makino H., Hwang E. J., Hedrick N. G., Komiyama T (2016) **Circuit Mechanisms of Sensorimotor Learning** *Neuron* **92**:705–721 https://doi.org/10.1016/j.neuron.2016.10.029

35 Romo R., Rossi-Pool R (2020) **Turning Touch into Perception** *Neuron* **105**:16–33 https://doi.org/10.1016/j.neuron.2019.11.033

36 Juavinett A. L., Erlich J. C., Churchland A. K (2018) **Decision-making behaviors: weighing ethology, complexity, and sensorimotor compatibility** *Curr Opin Neurobiol* **49**:42–50 https://doi.org/10.1016/j.conb.2017.11.001

37 Schoups A., Vogels R., Qian N., Orban G (2001) **Practising orientation identification improves orientation coding in V1 neurons** *Nature* **412**:549–553 https://doi.org/10.1038/35087601

38 Yan Y., et al. (2014) **Perceptual training continuously refines neuronal population codes in primary visual cortex** *Nat Neurosci* **17**:1380–1387 https://doi.org/10.1038/nn.3805

39 Xin Y., et al. (2019) **Sensory-to-Category Transformation via Dynamic Reorganization of Ensemble Structures in Mouse Auditory Cortex** *Neuron* **103**:909–921 https://doi.org/10.1016/j.neuron.2019.06.004

40 Makino H., Komiyama T (2015) **Learning enhances the relative impact of top-down processing in the visual cortex** *Nat Neurosci* **18**:1116–1122 https://doi.org/10.1038/nn.4061

41 Chadwick A., et al. (2023) **Learning shapes cortical dynamics to enhance integration of relevant sensory input** *Neuron* **111**:106–120 https://doi.org/10.1016/j.neuron.2022.10.001

42 Khan A. G., et al. (2018) **Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex** *Nat Neurosci* **21**:851–859 https://doi.org/10.1038/s41593-018-0143-z

43 Lomber S. G., Meredith M. A., Kral A (2010) **Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf** *Nat Neurosci* **13**:1421–1427 https://doi.org/10.1038/nn.2653

44 Shams L., Seitz A. R (2008) **Benefits of multisensory learning** *Trends Cogn Sci* **12**:411–417 https://doi.org/10.1016/j.tics.2008.07.006

45 Proulx M. J., Brown D. J., Pasqualotto A., Meijer P (2014) **Multisensory perceptual learning and sensory substitution** *Neurosci Biobehav Rev* **41**:16–25 https://doi.org/10.1016/j.neubiorev.2012.11.017

46 Komiyama T., et al. (2010) **Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice** *Nature* **464**:1182–1186 https://doi.org/10.1038/nature08897

47 Bimbard C., et al. (2023) **Behavioral origin of sound-evoked activity in mouse visual cortex** *Nat Neurosci* **26**:251–258 https://doi.org/10.1038/s41593-022-01227-x

48 Stringer C., et al. (2019) **Spontaneous behaviors drive multidimensional, brainwide activity** *Science* **364** https://doi.org/10.1126/science.aav7893

49 Niell C. M., Stryker M. P (2010) **Modulation of visual responses by behavioral state in mouse visual cortex** *Neuron* **65**:472–479 https://doi.org/10.1016/j.neuron.2010.01.033

50 Vinck M., Batista-Brito R., Knoblich U., Cardin J. A (2015) **Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding** *Neuron* **86**:740–754 https://doi.org/10.1016/j.neuron.2015.03.028

51 Musall S., Kaufman M. T., Juavinett A. L., Gluf S., Churchland A. K (2019) **Single-trial neural dynamics are dominated by richly varied movements** *Nat Neurosci* **22**:1677–1686 https://doi.org/10.1038/s41593-019-0502-4

52 Oude Lohuis M. N., Marchesi P., Olcese U., Pennartz C. M. A (2024) **Triple dissociation of visual, auditory and motor processing in mouse primary visual cortex** *Nat Neurosci* **27**:758–771 https://doi.org/10.1038/s41593-023-01564-5

53 Laurienti P. J., Kraft R. A., Maldjian J. A., Burdette J. H., Wallace M. T (2004) **Semantic congruence is a critical factor in multisensory behavioral performance** *Exp Brain Res* **158**:405–414 https://doi.org/10.1007/s00221-004-1913-2

54 Britten K. H., Shadlen M. N., Newsome W. T., Movshon J. A (1992) **The analysis of visual motion: a comparison of neuronal and psychophysical performance** *J Neurosci* **12**:4745–4765

Author information

Song Chang

Key Laboratory of Brain Functional Genomics (Ministry of Education and Shanghai), School of Life Sciences, East China Normal University, Shanghai, China

Beilin Zheng

College of Information Engineering, Hangzhou Vocational & Technical College, Hangzhou, China

Les Keniston
Department of Biomedical Sciences, Kentucky College of Osteopathic Medicine, University of Pikeville, Pikeville, United States

Jinghong Xu
Key Laboratory of Brain Functional Genomics (Ministry of Education and Shanghai), School of Life Sciences, East China Normal University, Shanghai, China
ORCID iD: 0000-0002-2864-4196

For correspondence: jhxu@bio.ecnu.edu.cn

Liping Yu
Key Laboratory of Brain Functional Genomics (Ministry of Education and Shanghai), School of Life Sciences, East China Normal University, Shanghai, China
ORCID iD: 0000-0002-9771-5971

For correspondence: lpyu@bio.ecnu.edu.cn

Editors

Reviewing Editor
Andrew King
University of Oxford, Oxford, United Kingdom

Senior Editor
Andrew King
University of Oxford, Oxford, United Kingdom

Reviewer #1 (Public review):

Summary:

Chang and colleagues used tetrode recordings in behaving rats to study how learning an audiovisual discrimination task shapes multisensory interactions in the auditory cortex. They found that a significant fraction of neurons in the auditory cortex responded to visual (crossmodal) and audiovisual stimuli. Both auditory-responsive and visually-responsive neurons preferentially responded to the cue signaling the contralateral choice in the two-alternative forced choice task. Importantly, multisensory interactions were similarly specific for the congruent audiovisual pairing for the contralateral side.

Strengths:

The experiments were conducted in a rigorous manner. Particularly thorough are the comparisons across cohorts of rats trained in a control task, in a unisensory auditory discrimination task, and the multisensory task, while also varying the recording hemisphere and behavioral state (engaged vs. anesthesia). The resulting contrasts strengthen the authors' findings and rule out important alternative explanations. Through the comparisons, they show that the enhancements of multisensory responses in the auditory cortex are specific to the paired audiovisual stimulus and specific to contralateral choices in correct trials and thus dependent on learned associations in a task-engaged state.

Weaknesses:

The main result is that multisensory interactions are specific for contralateral paired audiovisual stimuli, which is consistent across experiments and interpretable as a learned task-dependent effect. However, the alternative interpretation of behavioral signals is crucial to rule out, which would also be specific to contralateral, correct trials in trained animals. Although the authors focus on the first 150 ms after cue onset, some of the temporal profiles of activity suggest that choice-related activity could confound some of the results.

The auditory stimuli appear to be encoded by short transient activity (in line with much of what we know about the auditory system), likely with onset latencies (not reported) of 15-30 ms. Stimulus identity can be decoded (Figure 2j) apparently with an onset latency around 50-75 ms (only the difference between A and AV groups is reported) and can be decoded near perfectly for an extended time window, without a dip in decoding performance that is observed in the mean activity Figure 2e. The dynamics of the response of the example neurons presented in Figures 2c and d and the average in 2e therefore do not entirely match the population decoding profile in 2j. Population decoding uses the population activity distribution, rather than the mean, so this is not inherently problematic. It suggests however that the stimulus identity can be decoded from later (choice-related?) activity. The dynamics of the population decoding accuracy are in line with the dynamics one could expect based on choice-related activity. Also the results in Figures S2e,f suggest differences between the two learned stimuli can be in the late phase of the response, not in the early phase.

First, it would help to have the same time axis across panels 2,c,d,e,j,k. Second, a careful temporal dissociation of when the central result of multisensory enhancements occurs in time would discriminate better early sensory processing-related effects versus later decision-related modulations.

In the abstract, the authors mention "a unique integration model", "selective multisensory enhancement for specific auditory-visual pairings", and "using this distinct integrative mechanisms". I would strongly recommend that the authors try to phrase their results more concretely, which I believe would benefit many readers, i.e. selective how (which neurons) and specific for which pairings?

https://doi.org/10.7554/eLife.102926.1.sa2

Reviewer #2 (Public review):

Summary

In this study, rats were trained to discriminate auditory frequency and visual form/orientation for both unisensory and coherently presented AV stimuli. Recordings were made in the auditory cortex during behaviour and compared to those obtained in various control animals/conditions. The central finding is that AC neurons preferentially represent the contralateral-conditioned stimulus - for the main animal cohort this was a 10k tone and a vertically oriented bar. Over 1/3rd of neurons in AC were either AV/V/A+V and while a variety of multisensory neurons were recorded, the dominant response was excitation by the correctly oriented visual stimulus (interestingly this preference was absent in the visual-only neurons). Animals performing a simple version of the task in which responses were contingent on the presence of a stimulus rather than its identity showed a smaller proportion of AV stimuli and did not exhibit a preference for contralateral conditioned stimuli. The contralateral conditioned dominance was substantially less under anesthesia in the trained animals and was present in a cohort of animals trained with the reverse left/right contingency. Population decoding showed that visual cues did not increase the performance of the decoder but accelerated the rate at which it saturated. Rats trained on auditory and

then visual stimuli (rather than simultaneously with A/V/AV) showed many fewer integrative neurons.

Strengths

There is a lot that I like about this paper - the study is well-powered with multiple groups (free choice, reversed contingency, unisensory trained, anesthesia) which provides a lot of strength to their conclusions and there are many interesting details within the paper itself. Surprisingly few studies have attempted to address whether multisensory responses in the unisensory cortex contribute to behaviour - and the main one that attempted to address this question (Lemus et al., 2010, uncited by this study) showed that while present in AC, somatosensory responses did not appear to contribute to perception. The present manuscript suggests otherwise and critically does so in the context of a task in which animals exhibit a multisensory advantage (this was lacking in Lemus et al.,). The behaviour is robust, with AV stimuli eliciting superior performance to either auditory or visual unisensory stimuli (visual were slightly worse than auditory but both were well above chance).

Weaknesses

I have a number of points that in my opinion require clarification and I have suggestions for ways in which the paper could be strengthened. In addition to these points, I admit to being slightly baffled by the response latencies; while I am not an expert in the rat, usually in the early sensory cortex auditory responses are significantly faster than visual ones (mirroring the relative first spike latencies of A1 and V1 and the different transduction mechanisms in the cochlea and retina). Yet here, the latencies look identical - if I draw a line down the pdf on the population level responses the peak of the visual and auditory is indistinguishable. This makes me wonder whether these are not sensory responses - yet, they look sensory (very tightly stimulus-locked). Are these latencies a consequence of this being AuD and not A1, or ... ? Have the authors performed movement-triggered analysis to illustrate that these responses are not related to movement out of the central port, or is it possible that both sounds and visual stimuli elicit characteristic whisking movements? Lastly, has the latency of the signals been measured (i.e. you generate and play them out synchronously, but is it possible that there is a delay on the audio channel introduced by the amp, which in turn makes it appear as if the neural signals are synchronous? If the latter were the case I wouldn't see it as a problem as many studies use a temporal offset in order to give the best chance of aligning signals in the brain, but this is such an obvious difference from what we would expect in other species that it requires some sort of explanation.

Reaction times were faster in the AV condition - it would be of interest to know whether this acceleration is sufficient to violate a race model, given the arbitrary pairing of these stimuli. This would give some insight into whether the animals are really integrating the sensory information. It would also be good to clarify whether the reaction time is the time taken to leave the center port or respond at the peripheral one.

The manuscript is very vague about the origin or responses - are these in AuD, A1, AuV... ? Some attempts to separate out responses if possible by laminar depth and certainly by field are necessary. It is known from other species that multisensory responses are more numerous, and show greater behavioural modulation in non-primary areas (e.g. Atilgan et al., 2018).

https://doi.org/10.7554/eLife.102926.1.sa1

Reviewer #3 (Public review):

Summary:

The manuscript by Chang et al. aims to investigate how the behavioral relevance of auditory and visual stimuli influences the way in which the primary auditory cortex encodes auditory, visual, and audiovisual information. The main result is that behavioral training induces an increase in the encoding of auditory and visual information and in multisensory enhancement that is mainly related to the choice located contralaterally with respect to the recorded hemisphere.

Strengths:

The manuscript reports the results of an elegant and well-planned experiment meant to investigate if the auditory cortex encodes visual information and how learning shapes visual responsiveness in the auditory cortex. Analyses are typically well done and properly address the questions raised

Weaknesses:

Major

(1) The authors apparently primarily focus their analyses of sensory-evoked responses in approximately the first 100 ms following stimulus onset. Even if I could not find an indication of which precise temporal range the authors used for analysis in the manuscript, this is the range where sensory-evoked responses are shown to occur in the manuscript figures. While this is a reasonable range for auditory evoked responses, the same cannot be said for visual responses, which commonly peak around 100-120 ms, in V1. In fact, the latency and overall shape of visual responses are quite different from typical visual responses, that are commonly shown to display a delay of up to 100 ms with respect to auditory responses. All traces that the authors show, instead, display visual responses strikingly overlapping with auditory ones, which is not in line with what one would expect based on our physiological understanding of cortical visually-evoked responses. Similarly, the fact that the onset of decoding accuracy (Figure 2j) anticipates during multisensory compared to auditory-only trials is hard to reconcile with the fact that visual responses have a later onset latency compared to auditory ones. The authors thus need to provide unequivocal evidence that the results they observe are truly visual in origin. This is especially important in view of the ever-growing literature showing that sensory cortices encode signals representing spontaneous motor actions, but also other forms of non-sensory information that can be taken prima facie to be of sensory origin. This is a problem that only now we realize has affected a lot of early literature, especially - but not only - in the field of multisensory processing. It is thus imperative that the authors provide evidence supporting the true visual nature of the activity reported during auditory and multisensory conditions, in both trained, free-choice, and anesthetised conditions. This could for example be achieved causally (e.g. via optogenetics) to provide the strongest evidence about the visual nature of the reported results, but it's up to the authors to identify a viable solution. This also applies to the enhancement of matched stimuli, that could potentially be explained in terms of spontaneous motor activity and/or pre-motor influences. In the absence of this evidence, I would discourage the author from drawing any conclusion about the visual nature of the observed activity in the auditory cortex.

(2) The finding that AC neurons in trained mice preferentially respond - and enhance - auditory and visual responses pertaining to the contralateral choice is interesting, but the study does not show evidence for the functional relevance of this phenomenon. As has become more and more evident over the past few years (see e.g. the literature on mouse PPC), correlated neural activity is not an indication of functional role. Therefore, in the absence of causal evidence, the functional role of the reported AC correlates should not be overstated by the authors. My opinion is that, starting from the title, the authors need to much more carefully discuss the implications of their findings.

MINOR:

(1) The manuscript is lacking what pertains to the revised interpretation of most studies about audiovisual interactions in primary sensory cortices following the recent studies revealing that most of what was considered to be crossmodal actually reflects motor aspects. In particular, recent evidence suggests that sensory-induced spontaneous motor responses may have a surprisingly fast latency (within 40 ms; Clayton et al. 2024). Such responses might also underlie the contralaterally-tuned responses observed by the authors if one assumes that mice learn a stereotypical response that is primed by the upcoming goal-directed, learned response. Given that a full exploration of this issue would require high-speed tracking of orofacial and body motions, the authors should at least revise the discussion and the possible interpretation of their results not just on the basis of the literature, but after carefully revising the literature in view of the most recent findings, that challenge earlier interpretations of experimental results.

(2) The methods section is a bit lacking in details. For instance, information about the temporal window of analysis for sensory-evoked responses is lacking. Another example: for the spike sorting procedure, limited details are given about inclusion/exclusion criteria. This makes it hard to navigate the manuscript and fully understand the experimental paradigm. I would recommend critically revising and expanding the methods section.

https://doi.org/10.7554/eLife.102926.1.sa0