



# Neural Integration of Audiovisual Sensory Inputs in Macaque Amygdala and Adjacent Regions

Liang Shan<sup>1,2</sup> · Liu Yuan<sup>1,3</sup> · Bo Zhang<sup>1,7</sup> · Jian Ma<sup>1</sup> ·  
Xiao Xu<sup>1</sup> · Fei Gu<sup>3,4</sup> · Yi Jiang<sup>3,4,5</sup> · Ji Dai<sup>1,2,3,6</sup>

Received: 16 January 2023 / Accepted: 13 February 2023 / Published online: 15 March 2023

© Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences 2023

**Abstract** Integrating multisensory inputs to generate accurate perception and guide behavior is among the most critical functions of the brain. Subcortical regions such as the amygdala are involved in sensory processing including vision and audition, yet their roles in multisensory integration remain unclear. In this study, we systematically investigated the function of neurons in the amygdala and adjacent regions in integrating audiovisual sensory inputs using a semi-chronic multi-electrode array and multiple combinations of audiovisual stimuli. From a sample of 332 neurons, we showed the diverse response patterns to audiovisual stimuli and the neural characteristics of bimodal over unimodal modulation, which could be classified into four types with differentiated regional origins. Using the hierarchical clustering method, neurons were further clustered into five

groups and associated with different integrating functions and sub-regions. Finally, regions distinguishing congruent and incongruent bimodal sensory inputs were identified. Overall, visual processing dominates audiovisual integration in the amygdala and adjacent regions. Our findings shed new light on the neural mechanisms of multisensory integration in the primate brain.

**Keywords** Macaque · Amygdala · Multisensory · Audiovisual integration · Neural activity · Multichannel recording

## Introduction

In humans and other primate species, the efficient integration of information from different sensory modalities (e.g., vision and audition) is vital for survival and plays an important role in social communication. For example, the co-occurrence of a sound enhances the human detection sensitivity for low-intensity visual targets [1]. This phenomenon

---

Liang Shan, Liu Yuan, and Bo Zhang contributed equally to this work.

---

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s12264-023-01043-8>.

---

✉ Yi Jiang  
yijiang@psych.ac.cn

✉ Ji Dai  
ji.dai@siat.ac.cn

- <sup>1</sup> CAS Key Laboratory of Brain Connectome and Manipulation, the Brain Cognition and Brain Disease Institute (BCBDI), Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
- <sup>2</sup> Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Shenzhen 518055, China
- <sup>3</sup> University of Chinese Academy of Sciences, Beijing 100049, China

- <sup>4</sup> State Key Laboratory of Brain and Cognitive Science, CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
- <sup>5</sup> Chinese Institute for Brain Research, Beijing 102206, China
- <sup>6</sup> Shenzhen Technological Research Center for Primate Translational Medicine, Shenzhen 518055, China
- <sup>7</sup> Key Laboratory of Brain Science, Zunyi Medical University, Zunyi 563000, China

is called audiovisual (AV) integration, which is even present in human newborns [2]. In the past few decades, extensive studies have been carried out to unravel the underlying neural mechanisms. Human neuroimaging studies and monkey electrophysiological studies have suggested several candidate cortical regions for AV integration, which include the superior temporal sulcus [3–6], posterior parietal cortex [7], ventral intraparietal area [8, 9], and lateral intraparietal area [10]. In addition, subcortical nuclei such as the superior colliculus (SC) are also thought to play an indispensable role in AV integration [11, 12]. Since congruent AV stimuli (e.g., visual fear and auditory fear) accelerate the detection of fear [13] and it has long been hypothesized that the SC together with the pulvinar and amygdala constitute a subcortical pathway for fast fear responses [14], it is possible that the downstream nucleus of the SC—the amygdala—is also engaged in AV integration. Despite the evidence supporting a connection between the amygdala and the SC [15, 16], little is known regarding the function of the amygdala in AV integration.

Previous studies have shown that the amygdala is activated by visual looming stimuli in mice [17, 18]. Axons arriving from the basolateral amygdala to the auditory cortex are critically involved in the long-term retention of auditory fear memories [19]. In monkeys, neurons in the amygdala show distinct features in response to different facial expressions. Recently, amygdala neurons that respond to auditory, visual, or tactile stimuli have been identified [20]. However, it is not yet clear whether and how the amygdala contributes to multisensory processing. In addition, does the amygdala differ from transitional and adjacent regions such as the hippocampus and the pallidum in AV integration? Moreover, there are cases where the inputs from different sensory modalities are incongruent and even contradictory. In such a scenario, how the amygdala and adjacent regions integrate the incongruent information remains to be explored.

To answer these questions, a sufficient number of neurons should be characterized as these regions cover a relatively large area that consists of multiple subregions. Technically, in the past few decades, it has been a major challenge to record a large neuron population efficiently from a deep structure like the amygdala in primates. A semi-chronic microdrive [21] that enables multichannel and long-term recording, flexible adjustment of location, and ease of use has made high-quality single-unit recording from deep regions possible. On the other hand, as the data size increases with the number of channels, the complexity of data analysis also increases dramatically. To avoid unpredictable bias, introducing data-driven approaches such as machine learning to high-dimensional electrophysiological data analysis can boost the findings of unbiased results [22].

Therefore, we did not make any specific hypotheses regarding the functional differences between regions but drew conclusions from a data-driven approach.

We aimed to determine whether the amygdala and adjacent regions are actively involved in AV integration, whether there are specific sub-regions particularly responsible for AV integration, and how auditory and visual modalities interact. To this end, we recorded single-neuron activity from a large area around the amygdala in macaques using a semi-chronic multi-electrode array and a set of auditory, visual, or audiovisual stimuli. Through a combination of classical analytical methods with a data-driven approach (i.e., hierarchical clustering), our study provides new insight into the neural mechanisms of multisensory integration in the primate brain.

## Materials and Methods

### Animals

Two adult male macaques (Xingui Bio, Laibin, China) weighing 7.5 kg and 7 kg were used in the study. Both monkeys were housed in a primate facility with environmental control. The facility is accredited by The Association for Assessment and Accreditation of Laboratory Animal Care. All experimental procedures were approved by the Institutional Animal Care and Use Committee at Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, following the guidelines stated in the Guide for Care and Use of Laboratory Animals (Eighth Edition, 2011).

### Anesthesia and Surgeries

All surgical procedures were performed using sterile methods while the monkeys were anesthetized. For general anesthesia, the monkeys were first given atropine (0.05 mg/kg, intramuscular) to decrease bronchial secretions, and then ketamine (15 mg/kg, intramuscular) and propofol (6 mg/kg, i.v.) were given successively to induce and maintain anesthesia [23]. Electrocardiography, heart rate, oxygen saturation, and rectal temperature were continuously monitored (uMEC7, Mindray, Shenzhen, China).

The recording chamber (Form-fitting, PEEK) and the micro-drive (SC32-42mm, both from Gray Matter Research, Bozeman, USA) were implanted in a two-step procedure following the product manual. In brief, the recording chamber was implanted under the guidance of T1-weighted MRI (magnetic resonance imaging) images (3T Tim Trio scanner, Siemens, Munich, Germany). Then each monkey was run through a second MRI scan with a grid filled with Vitamin E to confirm the location of the chamber and register the

position of the electrodes. Two weeks after the first surgery, a craniotomy was performed and the micro-drive was installed vertically above the amygdala.

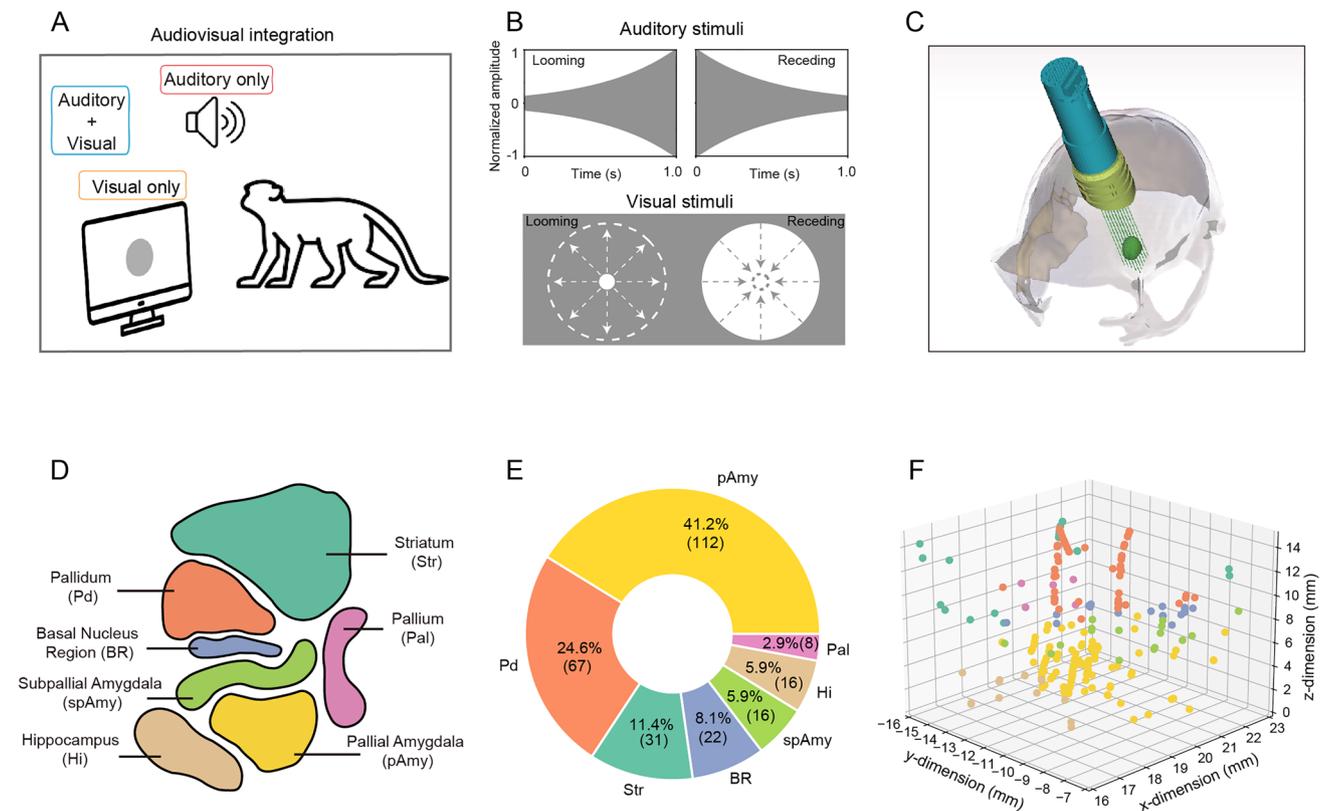
### Experimental Procedures and Electrophysiological Recording

During experiments, the monkey was trained to sit quietly in a chair with the head fixed. Visual stimuli were presented on a monitor (ASUS VG248, Taipei, China) 57 cm in front of the monkey, while auditory stimuli were presented from two speakers (Edifier R19U, Shenzhen, China) placed directly under the monitor. An eye tracker (iView X Hi-Speed Primate, SensoMotoric Instruments, Berlin, Germany) was used to monitor the eye position. The MatLab-based toolbox MonkeyLogic (National Institute of Mental Health, North Bethesda, USA) was used for experimental control.

During a recording session, the monkey was first required to fixate at the center of the screen for 500 ms, and then one of eight types of stimuli was presented: auditory looming (AL), auditory receding (AR), visual looming (VL), visual

receding (VR), auditory and visual looming (AL + VL), auditory looming and visual receding (AL + VR), auditory receding and visual looming (AR + VL), and auditory and visual receding (AR + VR). For visual stimuli, a disk presented at the center of the monitor gradually enlarged from 1° to 12° (looming, Fig. 1A, Video S1), or *vice versa* (receding, Video S2). For auditory stimuli, a 400-Hz complex tone composed of a triangular waveform was presented with intensity rising from 65 to 85 dB (looming, Fig. 1B, Audio S1), or *vice versa* (receding, Audio S2). Each stimulus was presented for 1 s. If a monkey successfully kept fixated during the presenting period it received a juice reward. In the meantime, the electrodes in the micro-drive (SC32, Gray Matter Research) were advanced carefully by tuning the bonded screws in steps of 1/4 to 1/2 turn (8 turns/mm).

A 128-channel electrophysiological recording system (OmniPlex, Plexon Inc., Dallas, USA) was used to monitor and record neural activity. Signals were filtered between 250 Hz and 5 kHz to record spiking activity. Spikes were sorted offline using Offline Sorter (Plexon Inc.) to identify single units. The peri-stimulus time histogram (PSTH) of the



**Fig. 1** Schematic of the experimental paradigm and the general results of electrophysiological recording. **A** Using the rhesus macaque as the subject, multisensory stimuli included auditory, visual, and a combination of auditory and visual inputs. **B** The auditory stimuli were a looming or receding tone, while the visual stimuli

were an expanding or shrinking disc, both lasting 1 s. **C** The micro-drive used for recording in the amygdala and adjacent regions. **D** The amygdala and adjacent regions identified in the current study. **E** The proportion of recorded neurons for each region. **F** The spatial distribution of recorded neurons for each region.

spike train for each neuron across different conditions (auditory only, visual only, and auditory + visual) was generated after aligning each trial to the stimulus onset and smoothed with the Savitzky-Golay filter (window length = 11, poly order = 2). The bin width was set to 25 ms.

### Electrode Localization

Electrode locations were determined by combining pre-operative anatomical MRI scans (acquired before implanting the microdrive), post-operative computed tomography (CT) scans, and tracking notes of electrode depths. After reorientation using SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/software/spm12>), CT and MRI images were co-registered using a rigid transformation algorithm in FieldTrip [24]. Then the electrode tracks were reconstructed by manually identifying two points—the electrode tip (the recording site) for each electrode—from the CT images. This procedure was facilitated by an interactive 3D scatter figure linked to the CT images [25]. Furthermore, the reconstructed electrode tracks were verified by overlaying them on the pre-operative MRI images with contrast fluid indicating the positions for lowering electrodes. The coordinates of each recording site in the native space were then calculated according to the notes that were kept of electrode depths across different recording sessions. The AFNI program @animal\_warper was used to align individual MRI images to the latest NIMH Macaque Template (NMT v2) [26], which provided non-linear transformation matrices between native and standard spaces. After transformation to native space, brain regions matching recording sites could therefore be obtained from the Subcortical Atlas of the Rhesus Macaque (SARM) [27]. Finally, to visualize and compare results at each recording site, the coordinates of each recording site were transformed into the standard space.

### Neural Responsivity to Bimodal Modalities

Neural responsivities to multisensory modalities were estimated using a sliding window nonparametric Wilcoxon rank-sum test [20]. In brief, rank-sum tests were used to compare the baseline-corrected mean firing rate between bimodal and the corresponding unimodal (i.e., AV vs A and AV vs V) in each 150-ms bin throughout the stimulus/post-stimulus period (in 50-ms bin steps). The alpha level was false discovery rate (FDR)-corrected for the number of comparisons (38 tests per trial) per set. Based on whether responses to AV were significantly different from A, V, or both A and V, neurons were classified as A-type, V-type, AV-type, and None. As we applied the statistical test for

AV vs A and AV vs V separately, a significance level of  $P < 0.0167$  was set for multiple comparison correction.

### AUC Analysis and the Hierarchical Clustering

The area under the receiver operating characteristic curve (AUC) [22] indicates how the firing rate can be discriminated from the baseline at any given time bin and is suitable for comparing the firing patterns of a large number of neurons over multiple events numerically as well as visually [28]. Here, AUC curves were calculated for each condition (AL + VL, AL + VR, AR + VL, AR + VR, AL, AR, VL, and VR) during the entire stimulus and post-stimulus period (bin = 100 ms, step = 25 ms). In brief, the histogram of the firing rate during the baseline period was compared with that during a given bin period by moving the criterion from zero to the maximum firing rate. Then the probability that the firing rate was greater than the criteria was extracted for the baseline and bin period, respectively, from which the receiver operating characteristic curve was generated by plotting these two probabilities as  $x$ - and  $y$ -axis [29, 30]. The area under this curve (the AUC) was then calculated to quantify the degree of overlap between these two firing rate distributions. Each AUC value lay between 0 and 1. A value  $< 0.5$  indicated a decrease in firing rating relative to the baseline, whereas a value  $> 0.5$  reflected an increase relative to the baseline.

The relative AUC was calculated between 8 pairs of bimodal and the corresponding unimodal stimuli throughout the stimulus and post-stimulus period [i.e., (AL + VL) – AL, (AL + VR) – AL, (AR + VL) – AR, (AR + VR) – AR, (AL + VL) – VL, (AR + VL) – VL, (AL + VR) – VR, (AR + VR) – VR]. Before the hierarchical clustering, the dimensional reduction was applied to the relative AUC profiles (displayed as a heatmap in Fig. 5) using independent component analysis [28]. Then, hierarchical clustering was applied using the standardized Euclidean distance metric and Ward's linkage method to the first 26 independent components of the relative AUC. The clustering threshold and cluster number were determined with the Silhouette score and the Davies-Bouldin index (Fig. S3).

### Results

During the experiments, the monkeys were required to fixate on the screen while an auditory stimulus (A), a visual stimulus (V), or a combination of auditory and visual stimuli (AV) was presented (Fig. 1A). The auditory stimulus was a looming (L) or receding (R) tone, while the visual stimulus was an enlarging or shrinking disc (Fig. 1B). By combining different sensory inputs, a stimulus set including eight conditions (AL, AR, VL, VR, AL + VL, AL + VR, AR

+ VL, and AR + VR) was established (see Audios S1, 2 and Videos S1, 2 for unimodal stimulus samples). For the bimodal conditions, auditory and visual stimuli were well aligned at onset and offset. We made extracellular recordings in the amygdala and adjacent regions using a semi-chronic 32-channel microdrive (Fig. 1C). By adjusting the depth of each electrode, we successfully identified 332 neurons and their locations. From a reconstruction of these recording sites using the SARM that provides 210 primary regions-of-interest [27], 272 neurons were registered to the amygdala or a peripheral region (Fig. 1D). These regions included the pallidum (Pd), striatum (Str), basal nucleus (BR), subpallial amygdala (spAmy), hippocampus (Hi), pallium, and pallial amygdala (pAmy). The labels of these regions were automatically generated by the SARM template, which reflects the subcortical parcellation and nomenclature of the 4th edition of The Rhesus Monkey Brain in Stereotaxic Coordinates [31]. The proportions of cells recorded from different regions are shown in Fig. 1E, of which 41.2% were recorded from the pAmy. Fig. 1F shows the spatial distribution of all 272 neurons with colors denoting brain regions (see Video S3 for 3D illustration).

### Diverse Neural Responses to Audiovisual Stimuli

First, to delineate how neurons in the amygdala and adjacent regions responded to simultaneously presented auditory and visual stimuli, we recorded the neural responses to bimodal (AV) stimuli and unimodal (A/V) stimuli by combining looming or receding conditions, and found diverse responses. For example, one neuron showed transient responses to auditory and audiovisual, but not visual stimuli, with short latency (Fig. 2A), while another responded transiently to all kinds of stimuli with different amplitudes (Fig. 2B). Of interest, one neuron showed increased responses to visual and audiovisual stimuli even after the stimuli had come to the end (Fig. 2C). Similarly, one cell showed sustained responses with a sharp onset to visual and audiovisual stimuli (Fig. 2D). While some neurons showed stronger responses to bimodal than to unimodal stimuli, we noted that the responses to audiovisual stimuli were evidently not the linear sum of responses to auditory and visual stimuli, let alone those neurons showing inhibitory responses to bimodal stimuli. Such diversity points out the complexity of neurons in the amygdala and adjacent regions in integrating multisensory information.

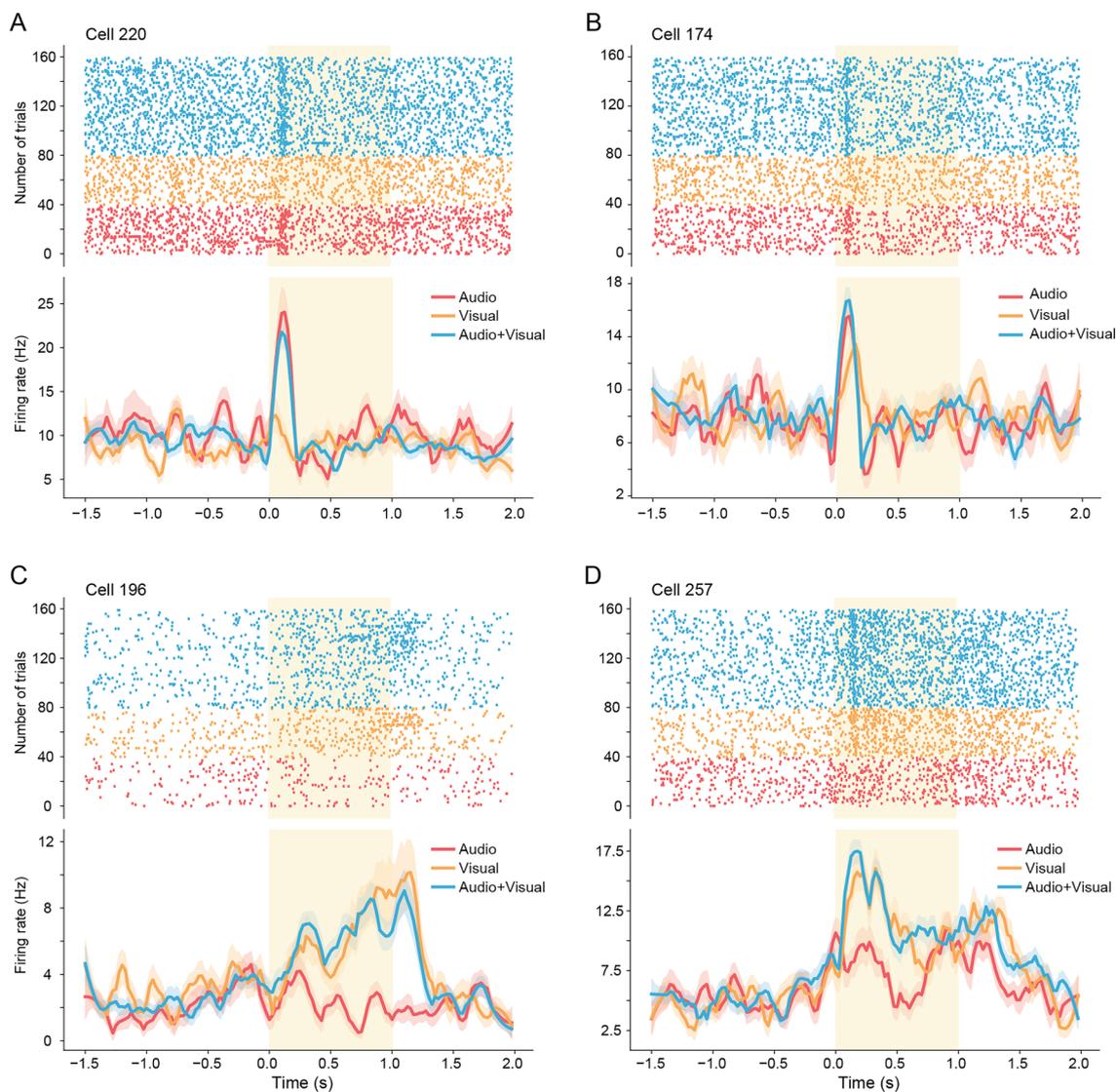
### Responses to Bimodal Versus Unimodal Stimuli

To characterize the response patterns to bimodal and unimodal stimuli, we plotted the mean response (baseline subtracted after pooling looming and receding conditions) to audiovisual stimuli *versus* the response to either the auditory

or the visual stimuli (Fig. 3A–D). During the stimulus period, as denoted by the confidence interval (CI, three times the SD), we noted that the distribution of audiovisual *vs* auditory stimuli was more dispersed than that of audiovisual *vs* visual stimuli (Fig. 3A, B). This disparity indicated that adding a visual to an auditory stimulus would likely change the auditory response ( $A + V \neq A$ , Fig. 3A). On the contrary, adding an auditory to a visual stimulus would induce hardly any change ( $A + V \approx V$ , Fig. 3B). Therefore, the visual component seems to dominate over the auditory component in terms of audiovisual integration or competition. When the same analysis was applied to the post-stimulus period, the difference between these two comparisons became much smaller (Fig. 3C, D), suggesting that the integration had come to the end at that time. We then subtracted the unimodal response from the bimodal response and plotted the differences for the stimulus and post-stimulus periods, in which the  $AV - A$  showed a wider distribution than  $AV - V$  (Fig. 3E, F). During the stimulus period, the means of  $AV - A$  and  $AV - V$  were both significantly greater than zero ( $P < 0.05$  and  $P < 0.001$ , respectively, *t*-test), indicating that, in general, bimodal stimuli induced stronger responses than unimodal stimuli. Since both positive and negative modulation occurred during bimodal compared to unimodal stimulations, to obtain the absolute modulation amplitude, we computed the absolute difference between bimodal and unimodal conditions ( $|AV - A|$  and  $|AV - V|$ ), which also demonstrated a stronger modulation for  $|AV - A|$  than  $|AV - V|$  during the stimulus period ( $P < 0.001$ , Wilcoxon signed-rank test, Fig. 3G) but not for the post-stimulus period (Fig. 3H). A further analysis using shorter time windows indicated that this difference between  $|AV - A|$  and  $|AV - V|$  was mainly reflected 200 ms after stimuli onset, as the initial 200 ms did not show equivalent differences (Fig. S1). The cross-regional analysis further indicated that the pAmy showed a significant attenuation for  $|AV - V|$  ( $P < 0.001$ ). Of note, the Hi seemed to show a marginal increase for  $|AV - V|$  ( $P = 0.06$ ), which may suggest a functional difference between these regions in audiovisual integration.

### Cell Classification Based on Bimodal Modulation

To further reveal the temporal dynamics of how bimodal stimuli modulate unimodal responses, we compared the bimodal responses with unimodal responses (also after pooling looming and receding conditions) using a sliding window. Nonparametric Wilcoxon rank-sum tests were used to find the period with significant differences. For example, the activity of the neurons shown in Fig. 4A–C show differentiated responses to AV than to A, V, and both A and V. Based on whether responses to AV were significantly different from A or V in any post-stimulus period ( $< 2$  s), we classified neurons into A-type (e.g., the

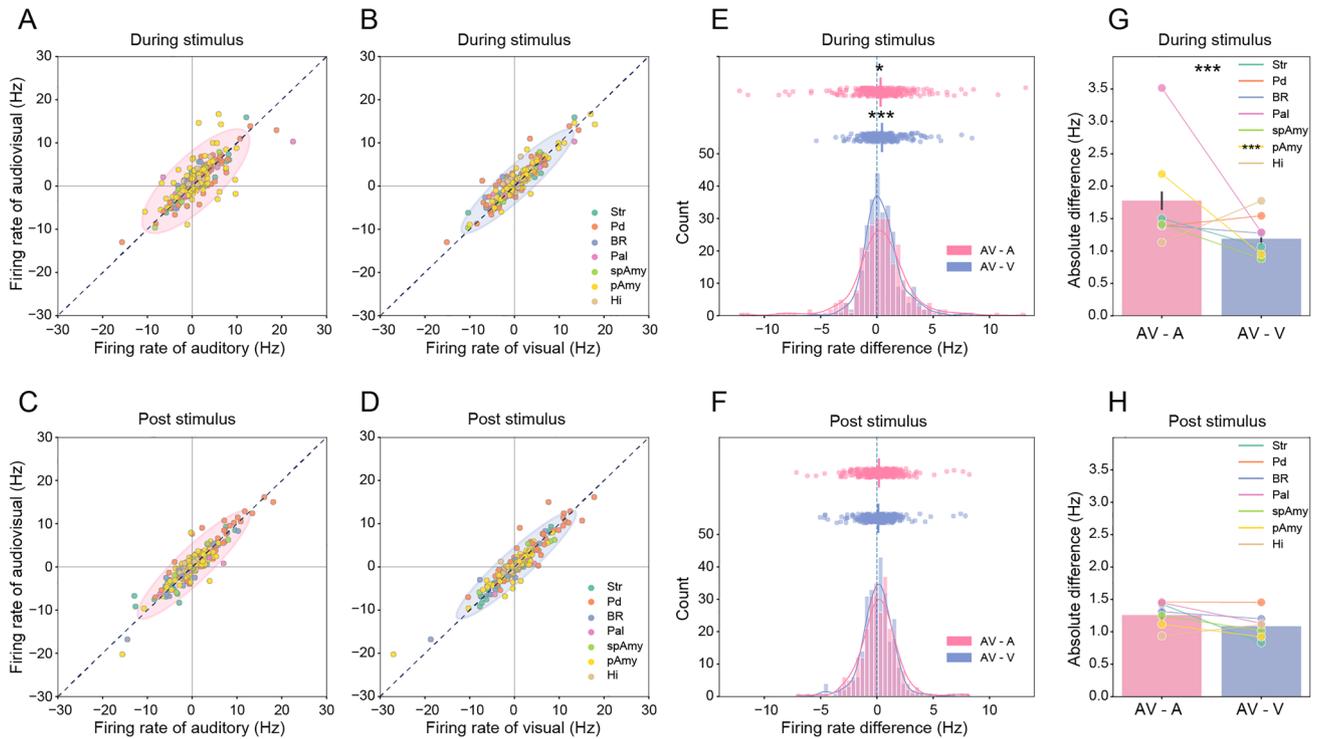


**Fig. 2** Examples of neurons that respond diversely to auditory, visual, and audiovisual stimuli. **A–D** Raster plots and spike density functions to audiovisual stimuli for four example neurons. Red, auditory;

yellow, visual; blue, audiovisual. The shaded rectangles indicate the stimulus period.

neuronal activity shown in Fig. 4A, 16.9%), V-type (e.g., the neuronal activity shown in Fig. 4B, 15.1%), AV-type (e.g., the neuronal activity shown in Fig. 4C, 40.1%), and None (27.9%, Fig. 4D, also see Fig. S2 for details). These cell types were found in all identified regions with different proportions. About half of the A-type, V-type, and None neurons originated from the pAmy, but only 30% (33/109) of AV-type neurons were from the pAmy, whereas 37% (40/109) were from the Pd. Of interest, most neurons in the Pd (60%, 40/67) were AV-type and only a small portion were other types. On the contrary, only 29% (33/112) of pAmy neurons were AV-type but 31% (35/112) were None. The statistics for A-, V-, and AV-types are

shown in Fig. 4E, in which the color-coded significant periods for each neuron are plotted against time and sorted in the order of the first significant point (the significant level was set to be  $P < 0.0167$  since multiple comparison correction was applied). Both positive (red) and negative (blue) modulation were found in all three cell types. It is worth noting that 68.3% of neurons were positively modulated by the bimodal stimuli (AV vs V) for the V-type, yet only 48.6% of neuronal visual responses were positively modulated by bimodal stimuli (AV vs V) for the AV-type. In general, for the AV vs V pairs, positive modulation occurred earlier than negative modulation. In addition, the AV vs A pairs seemed to show a wider distribution than the AV vs V pairs for the first 0.5 s. These data together



**Fig. 3** Comparisons of bimodal and unimodal responses. **A, B** Comparison of the audiovisual response during the stimulus period with the auditory and visual responses for each region. **C, D** Comparison of the audiovisual response during the post-stimulus period with the auditory and visual responses for each region. **E, F** Distribution of the response difference during stimulus period (**E**) and post-stimulus

period (**F**) between bimodal and unimodal conditions. Red, audiovisual subtracting auditory (AV – A); blue, audiovisual subtracting visual (AV – V). \* $P < 0.05$ ; \*\*\* $P < 0.001$  (one sample  $t$ -test,  $n = 272$ ). **G, H** The absolute difference during the stimulus period (**G**) and post-stimulus period (**H**) between bimodal and unimodal conditions. For **E–H**, \*\*\* $P < 0.001$  (Wilcoxon signed rank test,  $n = 272$ ).

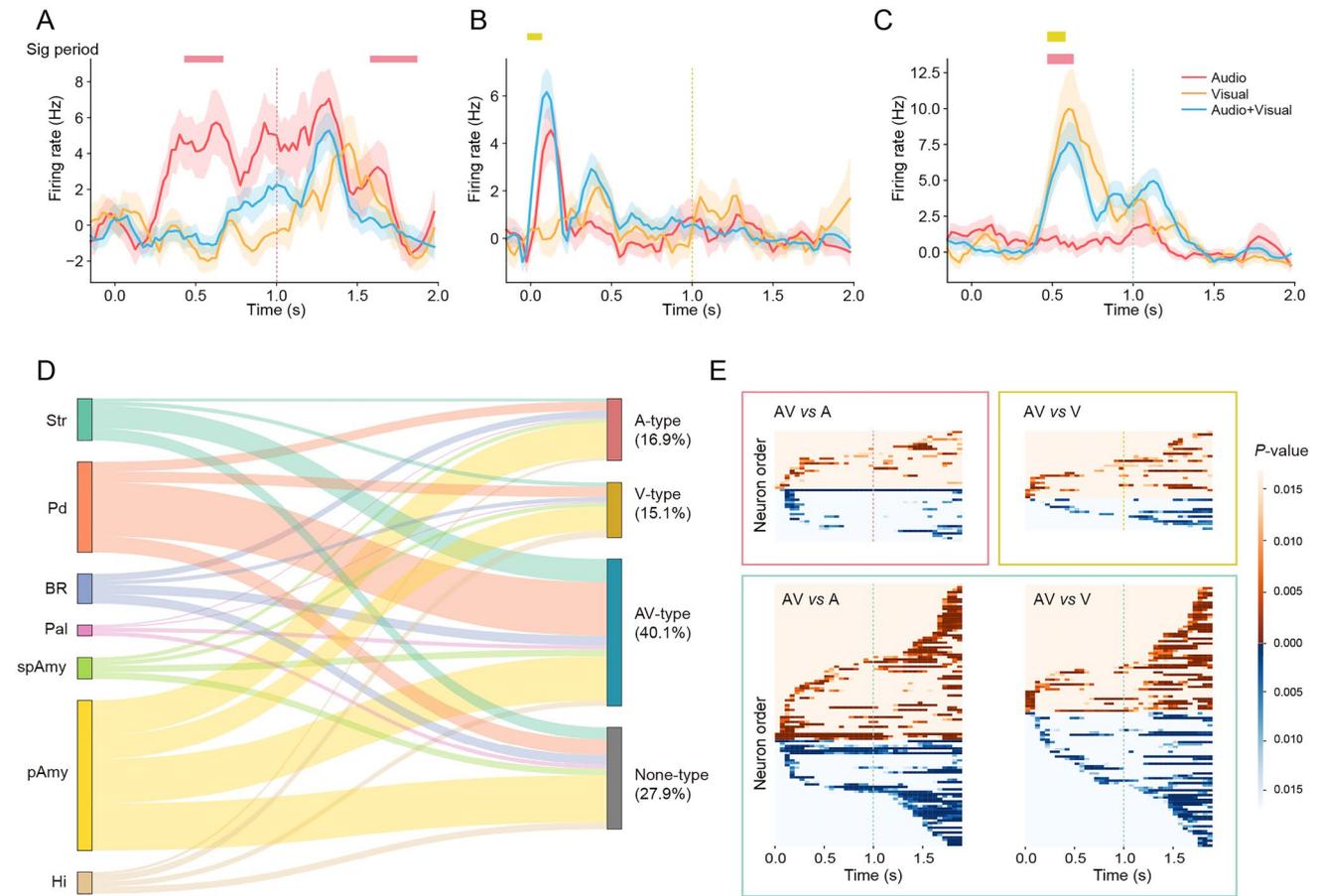
suggest that auditory and visual cues may take effect differentially on audiovisual integration.

### Cell Clustering Based on the Area under the Receiver Operating Characteristic Curve

The AUC [22] is an ideal indicator to discriminate a response from the baseline at any given time bin, thus enabling the simultaneous visualization of comparisons of the firing patterns of hundreds of neurons over multiple events. In addition, AUC-based hierarchical clustering offers the advantage of generating data-driven classifications regardless of prior knowledge. Therefore, to obtain an overview of the modulatory features of all modulated neurons (A-, V-, and AV-types,  $n = 192$ ), we first computed the AUC for each neuron and each condition (AL + VL, AL + VR, AR + VL, AR + VR, AL, AR, VL, and VR), and then obtained the relative AUC for eight pairs of bimodal conditions with subtraction of unimodal conditions [i.e., (AL + VL) – AL, (AL + VR) – AL, (AR + VL) – AR, (AR + VR) – AR, (AL + VL) – VL, (AR + VL) – VL, (AL + VR) – VR, and (AR + VR) – VR; see Methods for details]. Based on these relative AUCs, hierarchical clustering was applied to

all modulating neurons and the results are shown in Fig. 5 as a heat map (see Fig. S3 for threshold selection). Overall, we noted more warm zones than cold zones in the heat map, suggesting that bimodal input in general induced stronger responses than unimodal input, which was consistent with the results shown in Fig. 3E. These analyses showed that the modulating neurons could be clustered into five groups. Among them, we found two small subsets of neurons showing prominent stronger (Cluster 1) or weaker (Cluster 3) responses to AV than to A. The largest population (Cluster 2) showed stronger responses to AV than to A in general, and another large population (Cluster 4) showed stronger responses to V than to AV. The rest of the neurons were clustered into Cluster 5, which seemed to show complex modulation patterns.

Similarly, the mean of the relative AUC confirmed that the responses of neurons in Cluster 1 to AV showed the strongest positive modulation to auditory stimuli [especially (AR + VL) – AR], while those in Cluster 3 showed the strongest negative modulation instead (Fig. 6A), indicating that neurons in the amygdala and adjacent regions are able to discriminate different AV combinations. Also, we noted that the mean of the relative AUC showed a larger variation when



**Fig. 4** Neural types with bimodal over unimodal modulation. **A–C** Typical neurons show differentiated responses to AV than to A, V, and both A and V; the horizontal bars indicate the periods of significant difference. Pink bar,  $P < 0.0167$ ; yellow bar,  $P < 0.0167$ , Wilcoxon rank-sum test. **D** The proportions of A-type, V-type, AV-

type, and None, and their regional origins. **E** Temporal dynamics of the  $P$ -value for each neuron and each type. Red, positive modulation; blue, negative modulation. The color bar indicates the  $P$ -value for each neuron.

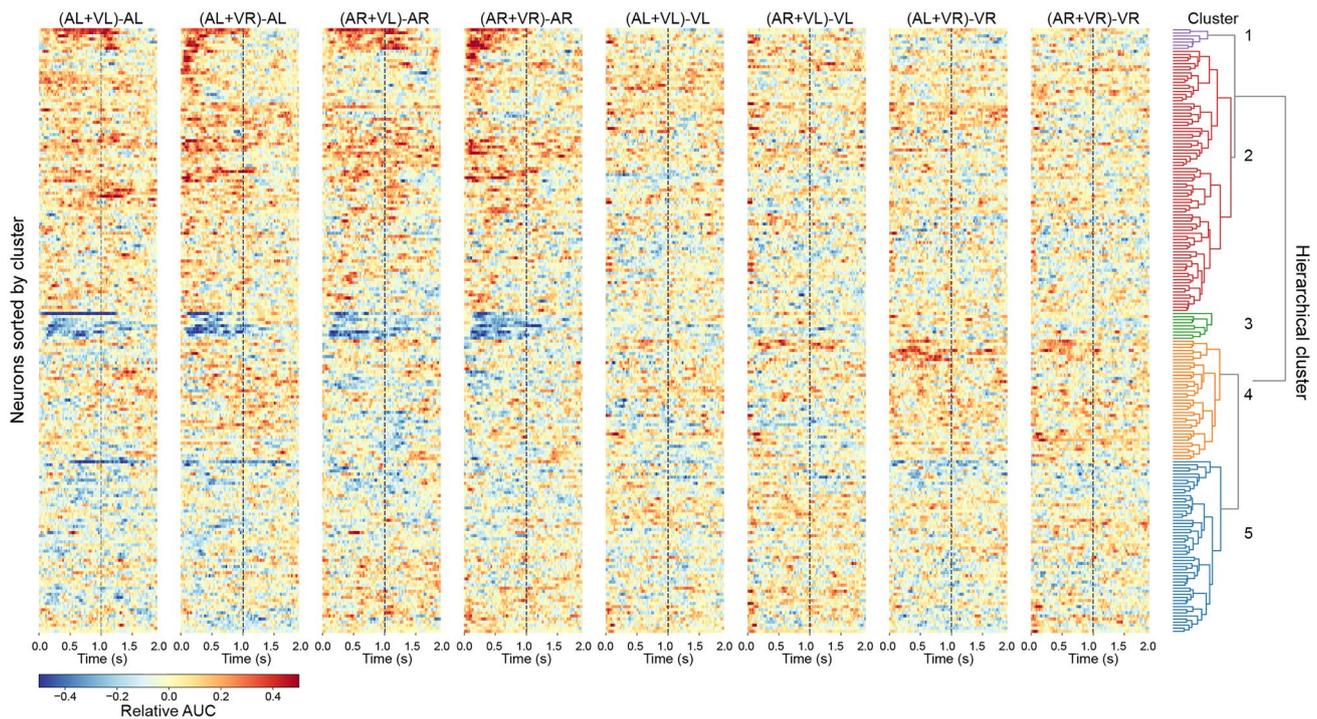
the auditory condition was used as the reference (Fig. 6A, left), whereas AV subtracting V remained mostly around zero (Fig. 6A, right), suggesting that visual stimuli are closer to audiovisual stimuli than auditory stimuli, which is also in line with the findings shown in Fig. 3.

The spatial distribution of these clusters is shown in Fig. 6B (see Video S4 for 3D animation), in which we noted that Clusters 1 and 3 seemed to localize in particular regions. Therefore, we analyzed the regional origins of all clusters and show the results as a Sankey plot in Fig. 6C. As expected, all Cluster 1 and most of Cluster 3 neurons were from the pAmy. Forty percent (34/85) of Cluster 2 neurons were also from the pAmy, which included the most diverse cell clusters. These together raised the prominence of pAmy in audiovisual integration. Within other regions, >68% (15/22) of Str and 54% (6/11) of Hi neurons were in Cluster 2, and 64% (7/11) of spAmy neurons were in Cluster 5, which was different from the pAmy (see also Fig. S2). Also, the functional association for all clusters (Fig. 6C,

right) showed that Cluster 3 was exceptional in that it did not include V-type neurons. Most Cluster 2 (55/85) neurons were AV-type, and only 13/85 neurons were V-type. About half of Cluster 4 (17/39) and Cluster 5 (28/56) neurons were AV-type (see also Fig. S2). These results together suggested functional differences between these clusters.

### Congruency Versus Incongruency in Bimodal Processing

Finally, we looked into how neurons in the amygdala and adjacent regions responded to congruent (AL + VL, AR + VR) and incongruent (AL + VR, AR + VL) bimodal sensory inputs. Here, AL + VL and AR + VR were defined as congruent as both sensory inputs generated the same perception of approaching or receding. By plotting the AUCs of congruent and incongruent conditions for all modulating neurons ( $n = 192$ , Fig. 7A), we found that, during the stimulus period, the AUCs of neurons in the pAmy and Hi showed



**Fig. 5** The relative AUC of each comparison for all modulating neurons. Neurons are sorted according to the clusters indicated on the right. The color bar indicates the value of the relative AUC.

larger deviations from 0.5 in the congruent condition than in the incongruent condition ( $P < 0.05$ ), whereas the BR showed a larger deviation to the incongruent condition ( $P < 0.05$ ). These data indicated that congruent stimuli induce stronger inhibitory responses than incongruent stimuli in the pAmy and Hi but have a reverse effect in the BR. However, during the post-stimulus period, no significant differences were found for any region between the congruent and incongruent conditions (Fig. 7B).

A similar analysis was applied to compare the looming and receding conditions (Fig. 7C, D), from which we found that the Pd and Str showed larger deviations in the receding condition during the stimulus period ( $P < 0.05$ ), while the spAmy showed larger deviation in the looming condition ( $P < 0.05$ ). These data may suggest that these regions have different sensitivities to looming and receding stimuli. Again, no difference was found for the post-stimulus period. However, due to the limited sample size of our current data set, these findings may need to be further examined in the future.

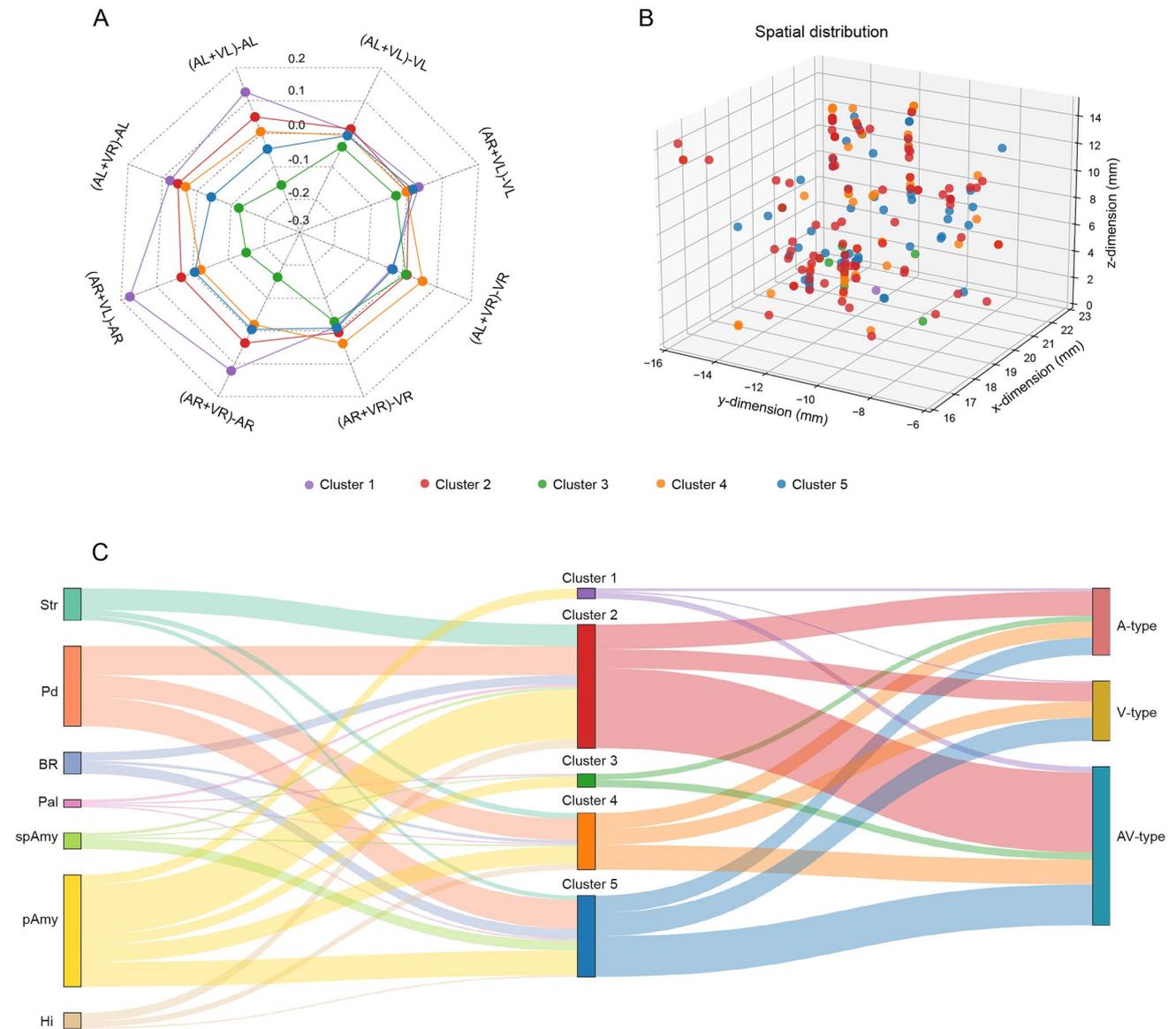
## Discussion

In the current study, through coarse to fine analyses, we tapped into the function of the amygdala and adjacent regions in integrating audiovisual sensory inputs from different perspectives. Specifically, following the illustration

of the experimental paradigm and example neurons (Figs 1 and 2), we first showed the average responses across the time domain for all neurons (Fig. 3), demonstrated the temporal dynamics after averaging responses to different sensory modalities (Fig. 4), illustrated the overall responses for all modulating neurons in both the time domain and the modality domain using data-driven approaches (Fig. 5), and then dissected the regional origins for different functional clusters (Fig. 6). In the end, we tried to delineate how neurons in these regions respond to congruency and incongruency between two modalities (Fig. 7). Overall, our findings widen current knowledge regarding the potential brain sites involved in multisensory integration from the cortex to the amygdala and adjacent regions, point out the dominance of visual input in audiovisual integration, and clarify the cell types and functional differences of the amygdala and adjacent regions, thus shedding new light on the neural mechanisms of multisensory integration in the primate brain.

## Audiovisual Integration in the Amygdala and Adjacent Regions

Previous studies have shown that multisensory integration occurs in the prefrontal cortex [32–34], parietal cortex [35, 36], superior temporal sulcus [37–42], and SC [41, 43]. In our study, by comparing bimodal responses with unimodal responses, we found that 40.1% of neurons

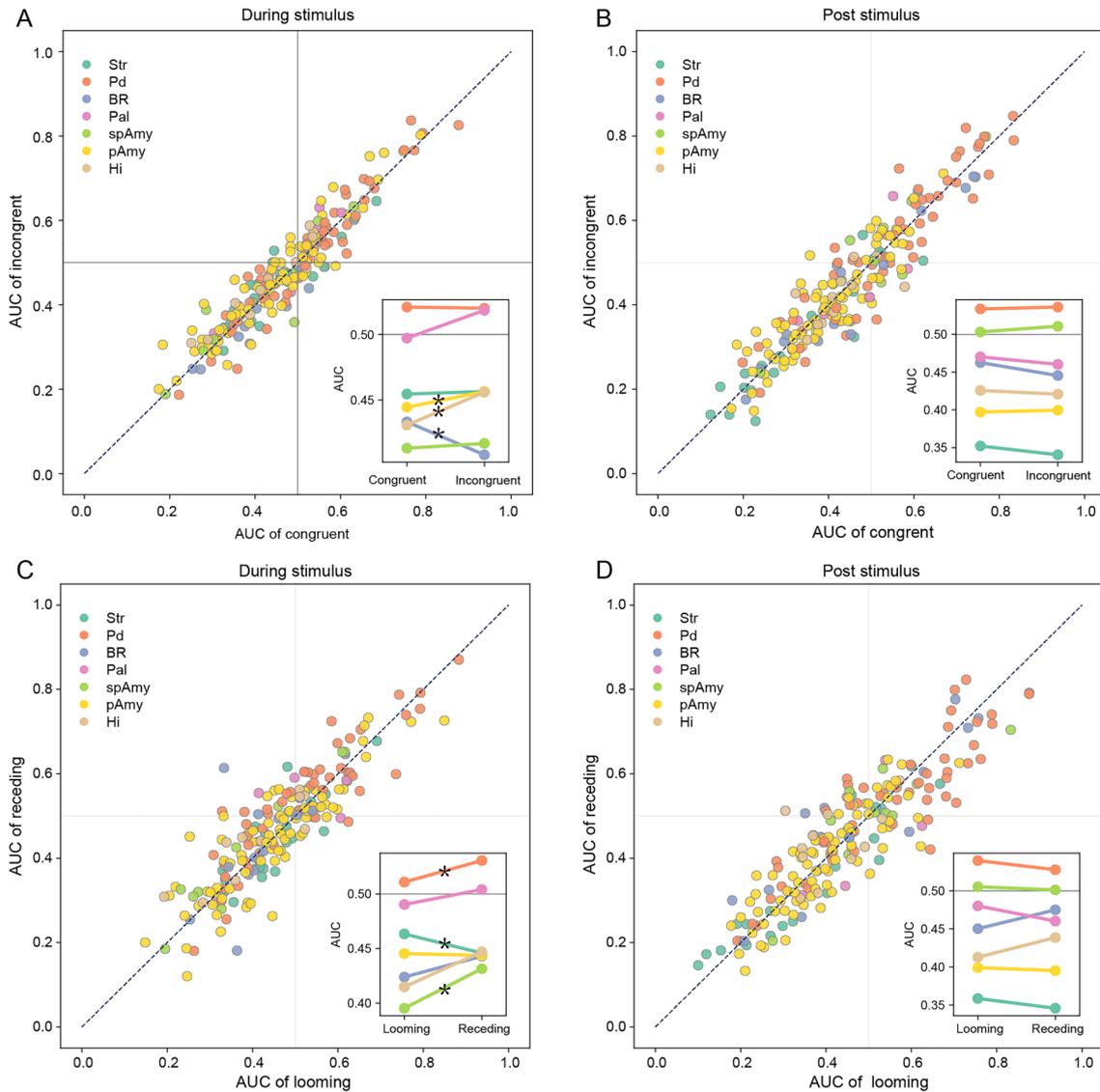


**Fig. 6** The characteristics of different clusters. **A** The mean AUC of each cluster for each comparison. **B** The spatial distribution of each cluster. **C** The regional origins and functions of each cluster.

(i.e. AV-type) responded differentially to AV than to both A and V. As the combination of A and V induced novel responses than responses to any component stimulus, these AV-type neurons were clearly multisensory. Therefore, in addition to the amygdala, its adjacent regions including the Str, Pd, and BR were all actively involved in audiovisual integration. This was not expected as these regions are normally believed to engage in movement control, reward processing, and mediating motivation. However, given that the ventral striatum receives connections from the orbitofrontal cortex and amygdala [44, 45], it is conceivable that these regions share similar properties in processing

multisensory input. In fact, the striatum together with the amygdala contains neurons that respond to taste or flavor [45–49]. Therefore, the amygdala and adjacent regions may work as a hub to integrate multisensory input including audition, vision, touch, and gustation.

Although different criteria were used, the proportion of multisensory neurons in our study (40.1%) seems to be comparable with a previous study, which found that 42.1% of neurons respond to at least two sensory modalities [20]. However, within these regions, the proportion of multisensory neurons in the pAmy (33/112) was much lower than in the Pd (40/67). Although this difference may arise from



**Fig. 7** Comparisons of congruent *versus* incongruent conditions and looming *versus* receding conditions. **A, B** Scatter plots of AUCs for congruent *versus* incongruent conditions during the stimulus period (**A**) and post-stimulus period (**B**). The inset indicates the mean AUC for each sub-region. **C, D** Scatter plots of AUCs for looming *versus*

receding conditions during the stimulus period (**C**) and post-stimulus period (**D**). \* $P < 0.05$ , Wilcoxon signed rank test,  $n = 22, 55, 15, 5, 11, 77,$  and  $11$  for the Str, Pd, BR, Pal, spAmy, pAmy, and Hi, respectively.

sampling bias, it still raises the possibility that the function of the Pd, especially the ventral Pd, has been underestimated in processing multisensory stimuli. Consistently, the ventral Pd has been shown to participate in gating sensorimotor behavior in rats [50–52]. Alternatively, as the ventral Pd also plays a role in mediating aversion, it is also possible that the looming and receding stimuli used in the current study aroused aversive responses in the Pd.

### Integration of Congruent and Incongruent Multisensory Inputs

In natural cases, animals must correctly interpret multisensory signals to guide behaviors. It is reasonable to assume that two congruent inputs from auditory and visual modalities should enhance the certainty of the interpretation, while two incongruent inputs should disrupt the interpretation. However, at the level of single neurons in our study, the neural responses did not seem to be precisely in line with such an assumption. Specifically, congruent auditory and visual stimuli did not evoke substantially greater responses than

incongruent stimuli, even though subtle differences were identified in the pAmy, Hi, and BR. This raises the possibility that the coding of certainty does not purely rely on the firing rate. For example, it may rely on temporal dynamics as well. As denoted in Fig. 4E, positive modulation of AV vs V occurred earlier than negative modulation. Similarly, in a previous study using stimulus decoding and information analysis, which incorporated response time course and response reliability, the responses of superior temporal sulcus neurons were found to convey more information during congruent audio-visual stimulation than incongruent bimodal stimulation [39]. Second, we found that visual input was the dominant modality when visual and auditory inputs were processed jointly. Therefore, adding auditory input to visual processing could not induce any drastic change, regardless of the congruency.

On the other hand, the stimulus set in the current study did not contain social information (e.g., facial expression, gender, and vocalization), which could potentially affect audiovisual integration as well. However, it is hard to infer the influence from current results. Therefore, it might be possible to develop a more advanced paradigm using virtual reality techniques to present multisensory stimuli with social cues included.

**Acknowledgements** We thank Dr. Yong-Shuai Ge for his kind assistance with the CT scans. This work was supported by the National Natural Science Foundation of China (U20A2017 and 31830037), Guangdong Basic and Applied Basic Research Foundation (2020A1515010785, 2020A1515111118, and 2022A1515010134), the Youth Innovation Promotion Association of the Chinese Academy of Sciences (2017120), the Shenzhen-Hong Kong Institute of Brain Science–Shenzhen Fundamental Research Institutions (NYKFKT2019009), Shenzhen Technological Research Center for Primate Translational Medicine (F-2021-Z99-504979), the Strategic Research Program of the Chinese Academy of Sciences (XDBS01030100 and XDB32010300), Scientific and Technological Innovation 2030 (2021ZD0204300), and the Fundamental Research Funds for the Central Universities.

**Conflict of interest** The authors declare that there are no conflicts of interest.

## References

- Noesselt T, Tyll S, Boehler CN, Budinger E, Heinze HJ, Driver J. Sound-induced enhancement of low-intensity vision: Multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *J Neurosci* 2010, 30: 13609–13623.
- Orioli G, Bremner AJ, Farroni T. Multisensory perception of looming and receding objects in human newborns. *Curr Biol* 2018, 28: R1294–R1295.
- Baumann O, Greenlee MW. Neural correlates of coherent audio-visual motion perception. *Cereb Cortex* 2007, 17: 1433–1443.
- Khandhadia AP, Murphy AP, Romanski LM, Bizley JK, Leopold DA. Audiovisual integration in macaque face patch neurons. *Curr Biol* 2021, 31: 1826–1835.e3.
- Lewis R, Noppeney U. Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *J Neurosci* 2010, 30: 12329–12339.
- von Saldern S, Noppeney U. Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. *J Neurosci* 2013, 33: 8841–8849.
- Brang D, Taich ZJ, Hillyard SA, Grabowecky M, Ramachandran VS. Parietal connectivity mediates multisensory facilitation. *Neuroimage* 2013, 78: 396–401.
- Kaminiaz A, Schlack A, Hoffmann KP, Lappe M, Bremmer F. Visual selectivity for heading in the macaque ventral intraparietal area. *J Neurophysiol* 2014, 112: 2470–2480.
- Schlack A, Sterbing-D'Angelo SJ, Hartung K, Hoffmann KP, Bremmer F. Multisensory space representations in the macaque ventral intraparietal area. *J Neurosci* 2005, 25: 4616–4625.
- Roitman JD, Shadlen MN. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 2002, 22: 9475–9489.
- Stein BE, Stanford TR. Multisensory integration: Current issues from the perspective of the single neuron. *Nat Rev Neurosci* 2008, 9: 255–266.
- Stein BE, Stanford TR, Rowland BA. Multisensory integration and the society for neuroscience: Then and now. *J Neurosci* 2020, 40: 3–11.
- Collignon O, Girard S, Gosselin F, Roy S, Saint-Amour D, Lassonde M. Audio-visual integration of emotion expression. *Brain Res* 2008, 1242: 126–135.
- Kragel PA, Čeko M, Theriault J, Chen D, Satpute AB, Wald LW, *et al.* A human colliculus-pulvinar-amygdala pathway encodes negative emotion. *Neuron* 2021, 109: 2404–2412.e5.
- Liu X, Huang H, Snutch TP, Cao P, Wang L, Wang F. The superior colliculus: Cell types, connectivity, and behavior. *Neurosci Bull* 2022, 38: 1519–1540.
- Rafal RD, Koller K, Bultitude JH, Mullins P, Ward R, Mitchell AS, *et al.* Connectivity between the superior colliculus and the amygdala in humans and macaque monkeys: Virtual dissection with probabilistic DTI tractography. *J Neurophysiol* 2015, 114: 1947–1962.
- Shang C, Liu Z, Chen Z, Shi Y, Wang Q, Liu S, *et al.* BRAIN CIRCUITS. A parvalbumin-positive excitatory visual pathway to trigger fear responses in mice. *Science* 2015, 348: 1472–1477.
- Wei P, Liu N, Zhang Z, Liu X, Tang Y, He X, *et al.* Processing of visually evoked innate fear by a non-canonical thalamic pathway. *Nat Commun* 2015, 6: 6756.
- Yang Y, Liu DQ, Huang W, Deng J, Sun Y, Zuo Y, *et al.* Selective synaptic remodeling of amygdalocortical connections associated with fear memory. *Nat Neurosci* 2016, 19: 1348–1355.
- Morrow J, Mosher C, Gothard K. Multisensory neurons in the primate amygdala. *J Neurosci* 2019, 39: 3663–3675.
- Dotson NM, Hoffman SJ, Goodell B, Gray CM. A large-scale semi-chronic microdrive recording system for non-human Primates. *Neuron* 2017, 96: 769–782.e2.
- Shan L, Huang H, Zhang Z, Wang Y, Gu F, Lu M, *et al.* Mapping the emergence of visual consciousness in the human brain via brain-wide intracranial electrophysiology. *Innovation (Camb)* 2022, 3: 100243.
- Zhang Z, Shan L, Wang Y, Li W, Jiang M, Liang F, *et al.* Primate preoptic neurons drive hypothermia and cold defense. *Innovation (Camb)* 2023, 4: 100358.
- Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011, 2011: 156869.

25. Stolck A, Griffin S, van der Meij R, Dewar C, Saez I, Lin JJ, *et al.* Integrated analysis of anatomical and electrophysiological human intracranial data. *Nat Protoc* 2018, 13: 1699–1723.
26. Jung B, Taylor PA, Seidlitz J, Sponheim C, Perkins P, Ungerleider LG, *et al.* A comprehensive macaque fMRI pipeline and hierarchical atlas. *Neuroimage* 2021, 235: 117997.
27. Hartig R, Glen D, Jung B, Logothetis NK, Paxinos G, Garza-Villarreal EA, *et al.* The subcortical atlas of the *Rhesus macaque* (SARM) for neuroimaging. *Neuroimage* 2021, 235: 117996.
28. Kremer Y, Flakowski J, Rohner C, Lüscher C. Context-dependent multiplexing by individual VTA dopamine neurons. *J Neurosci* 2020, 40: 7489–7509.
29. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 2012, 482: 85–88.
30. Keshavarzi S, Bracey EF, Faville RA, Campagner D, Tyson AL, Lenzi SC, *et al.* Multisensory coding of angular head velocity in the retrosplenial cortex. *Neuron* 2022, 110: 532–543.e9.
31. Paxinos G, Petrides MHE. *The Rhesus Monkey Brain in Stereotaxic Coordinates*. 4th ed. London: Academic Press, 2021.
32. Diehl MM, Romanski LM. Responses of prefrontal multisensory neurons to mismatching faces and vocalizations. *J Neurosci* 2014, 34: 11233–11243.
33. Romanski LM, Hwang J. Timing of audiovisual inputs to the prefrontal cortex and multisensory integration. *Neuroscience* 2012, 214: 36–48.
34. Sugihara T, Diltz MD, Averbeck BB, Romanski LM. Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J Neurosci* 2006, 26: 11138–11147.
35. Bremen P, Massoudi R, van Wanrooij MM, van Opstal AJ. Audiovisual integration in a redundant target paradigm: A comparison between *Rhesus macaque* and man. *Front Syst Neurosci* 2017, 11: 89.
36. Duhamel JR, Colby CL, Goldberg ME. Ventral intraparietal area of the macaque: Congruent visual and somatic response properties. *J Neurophysiol* 1998, 79: 126–136.
37. Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cogn Neurosci* 2005, 17: 377–391.
38. Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 2004, 41: 809–823.
39. Dahl CD, Logothetis NK, Kayser C. Modulation of visual responses in the superior temporal sulcus by audio-visual congruency. *Front Integr Neurosci* 2010, 4: 10.
40. Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 2005, 25: 5004–5012.
41. Stein BE, Stanford TR, Rowland BA. Development of multisensory integration from the perspective of the individual neuron. *Nat Rev Neurosci* 2014, 15: 520–535.
42. Werner S, Noppeney U. Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb Cortex* 2010, 20: 1829–1842.
43. Wallace MT, Stein BE. Development of multisensory neurons and multisensory integration in cat superior colliculus. *J Neurosci* 1997, 17: 2429–2444.
44. Haber SN, Knutson B. The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology* 2010, 35: 4–26.
45. Rolls ET. Emotion and decision-making explained: A précis. *Cortex* 2014, 59: 185–193.
46. Rolls ET. The texture and taste of food in the brain. *J Texture Stud* 2020, 51: 23–44.
47. Rolls ET, Thorpe SJ, Maddison SP. Responses of striatal neurons in the behaving monkey. 1. Head of the caudate nucleus. *Behav Brain Res* 1983, 7: 179–210.
48. Strait CE, Slezzer BJ, Hayden BY. Signatures of value comparison in ventral striatum neurons. *PLoS Biol* 2015, 13: e1002173.
49. Williams GV, Rolls ET, Leonard CM, Stern C. Neuronal responses in the ventral striatum of the behaving macaque. *Behav Brain Res* 1993, 55: 243–252.
50. Kodsi MH, Swerdlow NR. Prepulse inhibition in the rat is regulated by ventral and caudodorsal striato-pallidal circuitry. *Behav Neurosci* 1995, 109: 912–928.
51. Kodsi MH, Swerdlow NR. Ventral pallidal GABA-A receptors regulate prepulse inhibition of acoustic startle. *Brain Res* 1995, 684: 26–35.
52. Swerdlow NR, Braff DL, Geyer MA. GABAergic projection from nucleus accumbens to ventral pallidum mediates dopamine-induced sensorimotor gating deficits of acoustic startle in rats. *Brain Res* 1990, 532: 146–150.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.