Nonlinear visuoauditory integration in the mouse superior colliculus

Abbreviated title: Sensory integration in the superior colliculus

Shinya Ito[1](sito@ucsc.edu), Yufei Si[2], Alan M. Litke[1], David A. Feldheim[2*] (dfeldhei@ucsc.edu)

1. Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, 1156 High St. Santa Cruz, CA 95064, USA
2. The Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, 1156 High St. Santa Cruz, CA 95064, USA
Shinya Ito, Yufei Si, Alan M. Litke, and David A. Feldheim

Abstract

Sensory information from different modalities is processed in parallel, and then integrated in associative brain areas to improve object identification and the interpretation of sensory experiences. The Superior Colliculus (SC) is a midbrain structure that plays a critical role in integrating visual, auditory, and somatosensory input to assess saliency and promote action. Although the response properties of the individual SC neurons to visuoauditory stimuli have been characterized, little is known about the spatial and temporal dynamics of the integration at the population level. Here we recorded the response properties of SC neurons to spatially restricted visual and auditory stimuli using large-scale electrophysiology. We then created a general, population-level model that explains the spatial, temporal, and intensity requirements of stimuli needed for sensory integration. We found that the mouse SC contains topographically organized visual and auditory neurons that exhibit nonlinear multisensory integration. We show that nonlinear integration depends on properties of auditory but not visual stimuli. We also find that a heuristically derived nonlinear modulation function reveals conditions required for sensory integration that are consistent with previously proposed models of sensory integration such as spatial matching and the principle of inverse effectiveness.

Introduction

Integrating stimuli from different modalities that occur at the same location and time improves our ability to identify and respond to external events[1]. The superior colliculus (SC, homologous to the tectum in lower vertebrates) contains a topographic representation of visual space in its superficial layers (sSC) that is aligned with maps of auditory space in its deeper layers (dSC), and has been used as a model to elucidate general principles of how input from visual and auditory sources integrate[1–3]. For example, the barn owl optic tectum has been a model to elucidate the mechanisms used during development to align the visual and auditory spatial maps[4,5] as well as to determine the circuitry used to

modulate visuoauditory responses[6,7]. Studies primarily using anesthetized cats have shown that the mammalian dSC also contains neurons that respond to both visual and auditory stimuli (visuoauditory neurons)[8]. Some of these neurons exhibit superlinear multisensory enhancement—a type of multisensory integration (MSI)—whereby a neuron's response to a combination of stimuli is greater than the linear sum of the individual unimodal responses[8–14]. These studies, performed at the single neuron level, have determined three features of the stimuli that are required to induce superlinear multisensory enhancement: 1) 'Spatial matching', whereby both stimuli need to come from the same direction in the sensory field; 2) 'temporal matching', whereby the visual and auditory stimuli must overlap in time; and 3) 'inverse effectiveness', in which multisensory enhancement is weakened when the strength of the stimuli is increased[1,8].

While studies based on the owl tectum and cat SC have many advantages, these models are currently limited in their ability to manipulate genes and the activity of neurons, both of which can be used to determine the underlying circuitry that promotes sensory integration and elucidate the mechanisms by which it develops. The mouse is well suited for such analysis, but it is not known if the mouse SC contains visuoauditory neurons that exhibit MSI. Here we used high-density silicon probe recordings in awake mice, presented with simultaneous visual and auditory stimuli, to characterize the MSI properties of mouse SC neurons. We found that the receptive fields (RFs) of visual/auditory bimodal neurons in the dSC overlap and are organized into a topographic map of azimuthal space with some bimodal neurons exhibiting MSI. We also found that the auditory responses of these neurons have early and late temporal components, with only the late component being non-monotonic to the stimulus intensity. To determine the stimuli needed to invoke MSI and the properties of integration, we created a population-level model of the visuoauditory responses of simultaneously recorded individual neurons in which nonlinearity is given to neurons through a single modulation function that is shared by all the neurons in a dataset. The resulting modulation function quantifies the conditions needed for MSI and recapitulates the spatial matching requirements of the stimuli needed to invoke MSI, as well as the inverse effectiveness principle.

Results

The receptive fields of visual/auditory neurons in the deep SC are aligned and form a topographic map of space

In order to determine how visual and auditory information is organized in the mouse SC we used silicon probe electrophysiology to record the spiking activity of SC neurons in awake behaving CBA/CaJ mice in response to visual and auditory spot stimuli, and then determined their spatial RFs to each stimulus (Fig. 1a, b). We recorded a total of 1281 neurons from 8 mice. The spatial distribution and the fraction of cells that are responsive to visual stimuli, auditory stimuli, and both are shown in Fig. 1c, d. Consistent with what is known about the location of visual and auditory neurons in the mouse SC[15–19], we find that visually responsive neurons are located in both the superficial and deep SC while auditory responsive neurons are found only in the deep (≳ 400 μm) SC. Fig. 1e-g shows spatially matching visual and auditory RFs of an example neuron. When we plot the RF azimuth vs. SC anteroposterior location of the neurons, we find that both visually and auditory responsive neurons exhibit a strong correlation between their physical location and their RF location (Fig. 1h, i); the visual

map has a (77 ± 3) °/mm slope and (-1.6 ± 1.9) ° offset, while the auditory map has a (73 ± 8) °/mm slope and (10 ± 5) ° offset. This demonstrates that mice, like other mammals, have aligned visual and auditory maps of azimuthal space in the SC[4,18,19].

When we compared the visual and auditory RFs of the 62 bimodal neurons identified in our recordings, we found that (98.4 ± 1.6) % (61) of them have overlapping visual and auditory RFs. For (76 ± 5) % (47) of these neurons, the auditory RF is larger than the visual RF. The correlation coefficient between the visual RF centers and auditory RF centers of all visuoauditory neurons is 0.77 (Fig. 1j). These results demonstrate that the mouse dSC contains neurons that receive both visual and auditory information from the same azimuthal direction, and that they are organized as a topographic map.

Unimodal visual and early auditory SC responses are monotonic; late auditory SC responses can be non-monotonic

As a first step to understand how visual and auditory responses interact, we analyzed the temporal structures and intensity tuning of unimodal visual and auditory responses. Similar to what has been reported in the ferret[20], we found that auditory responses can be temporally segregated into early (< 20 ms) and late (> 20 ms) components (Fig. 2b, c; example rasters are in Supplementary Fig. 2), while visual responses occur only in the late time segment (Table 1). When we measured the intensity tuning curves of the early and late responses, we found that all of the visual (100% (178 / 178)) and nearly all ((99.7 ± 0.3) % (357 / 358)) of the early auditory responses are monotonic in nature, meaning the firing rate of the neuron increases with increasing stimulus intensity (Fig. 2e, h). However, the late auditory responses of some neurons ((11.4 ± 1.9) % (33 / 290)) were non-monotonic (Fig. 2f, i, red data points) and had significantly higher firing rates in response to low-intensity auditory stimuli than to a maximum-intensity (50 dB) stimulus. The temporal profiles of visual and auditory responses demonstrate that auditory signals arrive at the SC earlier than visual signals; therefore, if the visual and auditory inputs are detected by the retina and ears at the same time, the SC circuit may already be conditioned (modulated) by the early auditory processing (either in the SC or elsewhere) before the visual input arrives (Supplementary Fig. 1). This modulation could also explain the non-monotonicity of the late auditory responses.

Response type	Number of neurons	Percentage (%)
Visual early	1	0.08 ± 0.08
Visual late	178	13.9 ± 1.0
Auditory early	358	27.9 ± 1.3
Auditory late	290	22.6 ± 1.2
Auditory early & late	170	13.3 ± 0.9
Visual & auditory late	65	5.1 ± 0.6
Visual or auditory late	403	31.5 ± 1.3

(Used for modeling)		

Table1: Percentage of neurons that respond to visual and auditory unimodal stimuli.

Modeling the multisensory responses reveals an audition-dependent modulation structure

The above results are consistent with the hypothesis that auditory signals processed in the early time window influence nonlinear processing in the late time window. If this hypothesis is correct, nonlinear multisensory processing should depend on the properties of the auditory, but not the visual stimuli. To test this hypothesis, we used a maximum-likelihood fit analysis of models that include a simple sigmoidal response plus an auditory or visual stimulation-dependent modulation term and asked which model better fits the data. (Fig. 3a; see Methods)

In order to evaluate the goodness-of-fit of these models to the multisensory responses, we used the deviance per degree of freedom (D/DF; this value becomes asymptotically close to the reduced χ^2 in the limit of a large data sample (Wilk's theorem); a D/DF = 1 is a perfect fit, see Methods). We calculated the D/DF of a monotonic, non-modulated linear addition model (L-model) to be 1.72 ± 0.12 (n = 8 mice). When an audition-dependent modulation function is added to the model (AM-model), D/DF decreases to 1.173 ± 0.018. While we observed a marginal improvement in the number of neurons that are well-described by the model (59% in L-model to 72% in AM-model; $p < 0.01$; see Methods), we observed a dramatic reduction of the outlier deviance (D/DF - 1) in the AM model (0.72 to 0.173). This improvement is not seen if the modulation function is vision-dependent (VM-model; 1.66 ± 0.11; Fig. 3b). Therefore, we confirmed the above hypothesis that the nonlinear processing of multisensory responses can be better described by the modulatory properties of auditory compared to visual stimuli.

The modulation function of the model provides a quantitative description of the multisensory integration response properties for all neurons in the data set under a wide variety of auditory stimuli. Fig. 3c, d shows the average shape of the modulation function from 8 datasets (those for each dataset are shown in Supplementary Fig. 3) and has two important features. First, it has a bounded spatial distribution with a positive center and a negative surround. Second, more intense auditory stimuli cause stronger inhibition than less intense auditory stimuli. These two properties result in the modulation function being positive only near the location of the auditory stimulation, and only if the auditory stimulus is weak. The first property is known as spatial alignment of different sources (or surround suppression in other contexts[21,22]); the second property is known as the principle of inverse effectiveness[1]. These two previously described phenomena are now quantified with one modulation function, and our model-based analysis finds them heuristically, without prior knowledge of the system.

The AM-model improves the identification of neurons with multisensory processing

The utility of modeling is that it allows us to compress data from a high-dimensional stimulus space into a low-dimensional model parameter space. This overcomes the difficulty of analyzing and testing the significance of the neurons' response properties created by the large number (168) of visual and auditory stimulus patterns to identify those that exhibit multisensory interactions (see Materials and

Methods). For example, we used four variables—the locations and intensities of the visual and auditory stimuli—to make our stimuli (Fig. 3a), thus creating a four dimensional space and 168 different patterns of stimuli. The resulting data are not only difficult to visualize but also suffer from the statistical issues of multiple comparisons.

By using the AM-model to identify neurons that exhibit multisensory interactions we find that 47 % (190 / 403) of the modeled neurons exhibit significant modulation ($\alpha > 1$ with p < 0.01) including 17 % (70 / 403) of the neurons with superlinear addition ($\alpha^{\delta} > 1$ with p < 0.01; example: Fig. 4c,d, Supplementary Fig. 4) and 13 % (53 / 403) neurons with sublinear addition ($\alpha^{\delta} < 1$ with p < 0.01; example: Fig 4e,f) when the modulation function value (α^{δ}) is maximized. The distribution of maximum α^{δ} for significantly modulated neurons is shown in Fig. 4b.

Finding neurons that exhibit superlinear multisensory enhancement is difficult without a well-fit model; naively trying to find superlinear multisensory enhancement would require a significance test for all 144 combinations of the visuoauditory stimuli, and thus would suffer from a multiple comparison correction. For example, we found that only 4 neurons passed this significance test (p < 0.01 / 144). These results demonstrate the difficulty in the detection of multisensory enhancement using generic stimuli and how the model-based analysis overcomes this difficulty by connecting the population-level description of the modulation to an individual neuron's sensory integration.

Discussion

In this study, we present the first evidence of nonlinear multisensory integration (both enhancement and suppression) in the mouse SC, and demonstrate that the nonlinearity of the response depends on early auditory processing that modulates later visual/auditory responses. To achieve this, we used large-scale recording and a model-based analysis to quantify and visualize two principles of sensory integration that have been previously proposed based on studies of individual neurons: the spatial alignment of RFs and the principle of inverse effectiveness[1].

Topographic organization of visual, auditory, and bimodal neurons in the SC

The organization of sensory inputs into topographic maps and the subsequent integration of these maps in associative areas of the brain is a fundamental feature of neural processing. We confirm results from previous studies which show that the visual map of azimuthal space in the sSC is aligned with the auditory map of space in the dSC[17–19] and extend this finding to show that most of the visuoauditory bimodal neurons in the mouse SC have overlapping visual and auditory spatial RFs that themselves form a topographic map of space. This supports the model that topographic organization is used to align spatial visual and spatial auditory information perhaps to facilitate the integration of these two modalities.

Multisensory enhancement and inhibition in the mouse SC

We find that the mouse SC contains a significant percentage of stimulus-responsive neurons that exhibit superlinear multisensory enhancement (~17%) and sublinear addition (~13%; Fig 4b). A similar distribution of nonlinearity has been reported in the barn owl tectum[13]. In some previous studies the responses of single neurons are characterized by tailored stimuli (meaning that the time delay

between the visual and auditory stimuli were adjusted for each neuron) to increase the yield of multisensory integration. Instead, we performed large scale recordings that presented the same stimuli to all neurons as happens in the real world. This approach requires many different patterns of stimuli and to prevent the data from being sacrificed by a multiple comparison correction, we used a model-based approach. As large-scale recording methods are becoming more common[23–25], we believe that similar model based approaches should become the norm instead of using an ad-hoc index (such as multisensory index[26] or orientation selectivity index[16,27]) for each application.

Early auditory processing modulates the visuoauditory response

We present multiple results that suggest that early auditory processing induces visuoauditory integration in the SC: (1) only auditory responses arise in the early (< 20 ms) time window; (2) in the late (> 20 ms) time window, auditory responses are non-monotonic, while visual responses are monotonic; and (3) nonlinearity of the multisensory integration depends on the properties of auditory stimuli, but not visual stimuli. However, we did not identify the source of this modulation in this study. Potential sources include local circuitry within the SC[21,28], a pathway with larger latency such as a cortical pathway[29], and/or reciprocal connections with other brain areas such as the parabigeminal nucleus[30–32].

Conclusion

The auditory and visual RFs of the neurons in the mouse SC are overlapping and organized as a spatial topographic map and visuoauditory neurons exhibit nonlinear MSI. A model that extends a simple sigmoidal response with an audition-dependent modulation function can explain 76% of the outlier deviance of the simple sigmoidal model. The resulting modulation function is consistent with two important principles of multisensory integration: spatial matching and inverse effectiveness. These results open the door to further studies designed to determine the circuitry underlying multisensory integration, and the mechanisms used for its development.

Materials and Methods

Animal Preparation for electrophysiology

All procedures were performed in accordance with the University of California, Santa Cruz (UCSC) Institutional Animal Care and Use Committee.

Animal preparation, head plate attachment and craniotomy were performed as previously described[17]. We used 2–9 month-old CBA/CaJ (The Jackson Laboratory, 000654) mice of each sex. One to four days before the recording, a custom-made titanium head plate was attached to the mouse's skull. On the day of the recording, the mouse was anesthetized with isoflurane (3% induction, 1.5–2% maintenance; in 100% oxygen) and a craniotomy was made (~1.5–2 mm diameter) in the left hemisphere above the SC (0.6 mm lateral from the midline, on the lambdoid suture). A 256-channel silicon probe (provided by Prof. Masmanidis[25]) was inserted through the cortex into the SC with its four shanks aligned along the A–P axis (Fig. 1a). The probe was lowered until robust multi-unit visual responses from the superficial SC reached the top of the active region of the most posterior shank. The

probe was then retracted by 120–200 μm from its location to reduce drifting during the recording. The mice were euthanized after the recording session.

During the recording, the mouse was allowed to run freely on a cylindrical treadmill made of polystyrene foam. The movement of the cylinder was recorded by a shaft encoder (US Digital, H5-100-NE-S).

Stimulation

Visual stimulation

To measure visual receptive fields (RFs), a 10°-diameter flashing circular spot on a 16 × 7 grid with 10° spacing[16] was presented from a computer monitor (Samsung S29E790C; 67.3 cm x 23.4 cm active display size; refresh rate: 60 Hz; gamma corrected) placed 25 cm away from the animal. The stimuli were presented using a custom Matlab program that uses the Psychophysics Toolbox extensions[33–36]. The 500-ms flashes were either ON (white) or OFF (black) on a gray background at mean luminance (32 cd/m^2) and a 500-ms gray screen was inserted after each stimulus presentation. The stimulus was presented in a random order of the contrast and location. Every pattern was repeated 12 times. To estimate the RF parameters, a 2-dimensional Gaussian was fit to the collected neural responses (unbinned maximum-likelihood estimation[16]).

To collect the local field potentials (LFPs) used to estimate the surface location of the SC[16], a contrast-alternating checkerboard stimulus (0.04 cpd square wave alternating at 0.5 Hz) was presented.

Auditory stimulation

Virtual auditory space stimulation for mice was done in an anechoic chamber as previously described[17]. Briefly, sound is produced through earphones that are placed near the mouse's pinnae. The incident direction of the sound was controlled using filters derived from the head-related transfer functions of the mouse.

To measure the auditory receptive fields of the neurons, a full-field white-noise burst was presented from each virtual location in a grid space of 5 elevations (0° to 80° with 20° steps) and 17 azimuths (-144° to 144° with 18° steps), totaling 85 points in a two-dimensional directional space. A maximum-likelihood fit of the Kent distribution was used to determine the spatial properties of the auditory receptive fields[17]. Because the speakers for auditory stimulation blocked a part of the visual field where the azimuth is larger than 90°, we excluded neurons with RF azimuth larger than 90° (either visual or auditory) from the analysis.

Multisensory stimulation

To create a multisensory stimulation system the above visual and auditory stimulation conditions were combined with the visual stimulation computer used as a master computer for synchronization. To synchronize the timing of the stimuli, the visual stimulation computer sends a transistor-to-transistor logic (TTL) pulse to the auditory stimulation computer, and the auditory stimulation computer sends a TTL pulse to the data acquisition system (RHD2000, Intan Technologies). Because the entire system

did not fit in our anechoic chamber, this experiment was done without an anechoic chamber after verifying that the auditory RFs of the SC neurons are not significantly altered by removing the anechoic chamber (data not shown).

Visuoauditory spot stimulation was used to measure the properties of multisensory processing. The duration of the visual and auditory stimuli were 100 ms. The auditory stimulus is delivered ~13 ms later than the visual stimulus, mimicking the effect of the time delay due to the slow propagation speed of sound in the air (343 m/s; compared to the speed of light). This time delay is consistent with a simultaneous visual and auditory event occurring at 4–5 meters away from the animal. Temporal responses of neurons are calculated relative to the onset of the auditory stimulation. The intensity of the black spots were (0, -0.29, -0.50, -0.65, -0.75) in two of the experiments, and (0, -0.1, -0.2, -0.4, -0.8) in the other 6 experiments, measured as a fractional change from the gray background (for example, -0.1 is 10% reduction of the brightness level, or 28.8 cd/m^2 compared to the 32 cd/m^2 background). The intensity of the auditory stimuli (white noise bursts) were 20, 30, 40, and 50 dB SPL (within 5–80 kHz), or no sound. The stimuli were presented at three locations, 30 degrees apart, that match the visual RF positions of 3 of the silicon probe shanks. We presented black spots because black spots excite more SC neurons than white spots[37]. The total number of the patterns presented to the animal are 5 (visual intensities) × 5 (auditory intensities) × 3 (visual locations) × 3 (auditory locations) = 225 patterns. These patterns were presented in a random order in a 2-s period and repeated 30 times. When either no visual or auditory stimulus is given, this stimulus becomes unimodal, and is used for intensity tuning analysis, as described in the section on “Visual and early auditory responses are monotonic; late auditory responses can be non-monotonic”.

Data analysis

Blind analysis

We used blind analysis, an effective method for reducing false-positive reporting of results[38,39]. Of the 8 recording data sets, 4 were used for exploratory purposes and the other 4 were blinded. Once the features and parameters to be analyzed were decided the same analysis was done on the blinded dataset to confirm if the results agreed with those of the exploratory dataset. All of the results reported in the present study passed a significance test both in the exploratory dataset and the blinded dataset. The combined dataset was used to derive the p-values. Of the individual datasets shown in Supplementary Fig. 3, data 1-4 are exploratory datasets and data 5-8 are blinded datasets.

Spike-sorting

Spikes were sorted using custom-designed software[16,17,40]. Briefly, raw analog signals were high-pass filtered (cutoff ~313 Hz, wavelet filter[41]) and thresholded for spike detection. Detected spikes were clustered employing principal component analysis to its waveforms on the seed electrode and its surrounding electrodes. A mixture-of-Gaussians model was used for cluster identification. Clusters with many spikes in the refractory period and duplicated clusters for the same neuron (determined by isolation distance, L-ratio, cross-correlation analysis and similarity of the waveform shapes) were excluded from the analysis. Neurons with axonal waveforms were identified by principal component analysis and excluded from the analysis[17].

Estimation of the relative position of the neurons across datasets

To make a map of visual/auditory space across datasets from multiple mice, the relative physical locations of the SC neurons were estimated using the visual RFs. The visual RFs on multiple shanks were used to estimate the A–P position where the visual RF azimuth was 0°, and this point was defined as the zero point of the A–P SC position. The height of the SC surface is estimated by LFPs in response to a checkerboard visual stimulus[16]. Estimating that the insertion angle of the probe into the SC was 25°[17], the A–P positions and the depths of the neurons across multiple datasets were calculated and superimposed. Only neurons with a positive A–P position were analyzed in order not to include neurons outside the SC.

Model of multisensory responses

Our model comprises commonly used sigmoidal intensity tuning curves and an auditory-dependent modulation term derived from the analysis of the multisensory responses (AM-model, see Results). We formulated our multisensory model as follows.

For each neuron, the response R to a visuoauditory stimulus is determined by the following equations.

$$R(\varphi_V, \varphi_A, I_V, I_A) = \alpha^{\delta(\varphi_A, I_A)} (F_V(\varphi_V) S(I_V; I_{V50}, n_V) + F_A(\varphi_A) S(I_A; I_{A50}, n_A) + b)$$

where,

$$\delta(\varphi_A, I_A) = A_{Exc} M_{Exc}(\varphi_A, I_A) - A_{Inh} M_{Inh}(\varphi_A, I_A),$$

$$S(I; I_{50}, n) = \frac{I^n}{I_{50}{}^n + I^n},$$

and,

$$M_T(\varphi_A, I_A) = G(\varphi_{eff} - \varphi_A, \sigma_T)\, S(I_A; I_{T50}, n_T) .$$

$R(\varphi_V, \varphi_A, I_V, I_A)$ is the resulting firing rate of a neuron when a visual stimulus is presented at the azimuth φ_V with the intensity I_V and an auditory stimulus is presented at the azimuth φ_A with the intensity I_A; $G(\mu, \sigma)$ is a Gaussian with a unit height, mean μ and standard deviation σ; T is a sign of the modulation (Exc (excitatory) or Inh (inhibitory)), φ_{eff} is the neuron's effective RF azimuth (we used this instead of measured RF azimuths because some neurons do not have a measured RF azimuth); and σ is the width of the modulation effect along the azimuthal axis. The baseline (spontaneous) firing rate b is a fixed value that is measured from all the intervals of the stimuli and is not a model parameter. The descriptions and the ranges of the model parameters are summarized in Table 2. Parameters labelled as "Individual" apply independently to each neuron; Global parameters apply to all the neurons in a given dataset.

To test the model without modulation (L-model), we set $\alpha = 1$. By setting this value, the number of the global parameters go down from 8 to 0, and the number of the individual parameters go down from 12 to 10 (α and φ_{eff} become irrelevant). To test vision-dependent modulation (instead of auditory-dependent modulation; VM-model), we changed the power of the modulation term to the following.

$$\delta(\varphi_V, I_V) = A_{Exc} M_{Exc}(\varphi_V, I_V) - A_{Inh} M_{Inh}(\varphi_V, I_V)$$

Use of the parameters	Number of parameters	Individual or global	Range (Phase 1)	Range (Phase 2)
$F_V(\varphi_V)$: Max visually evoked firing rates	3 (one per stim location)	Individual	(0-100) Hz	(-100-200) Hz
I_{V50}, n_V : Visual intensity response sigmoid parameters	2	Individual	c_{V50} : 0.01-1 n_V : 0-5	c_{V50} : 0.01-3 n: 0-30
$F_A(\varphi_A)$: Max auditory evoked firing rates	3 (one per stim location)	Individual	(0-100) Hz	(-100-200) Hz
I_{A50}, n_A : Auditory intensity response sigmoid parameters	2	Individual	c_{A50} : 0.1-100 n_A : 0-5	c_{A50} : 0.1-300 n_A : 0-30
α : Strength of modulation	1	Individual	1.01-30	1-50
φ_{eff} : effective RF azimuth of the neuron	1	Individual	A fixed value estimated from the neuron's physical location	(-100-150) °
σ_T : Width of modulation	2 (excitatory & inhibitory)	Global	(18-100) °	(18-100) °
I_{T50}, n_T : Intensity response sigmoid for modulation	4 (excitatory & inhibitory)	Global	c_{T50} : 0.1-100 n_T : 0.1-5	c_{T50} : 0.1-200 n_T : 0.1-5
A_T : Relative amplitude of modulation	2 (excitatory & inhibitory)	Global	Exc: 0-10 Inh: 0-10	Exc: 0-20 Inh: 0-20
Total number of parameters	12 * N + 8			

Table 2: Summary of the model parameters

Although the total number of stimulus patterns are 225 as noted above, some of the patterns are redundant. Namely, when either visual or auditory stimulus intensity is zero, its location becomes an irrelevant variable. Thus, the number of non-redundant stimulus patterns are 144 (4 x 4 x 3 x 3; multisensory stimuli) + 12 (4 x 3, no visual stimuli) + 12 (4 x 3, no auditory stimuli) = 168. Therefore, the

number of degrees of freedom is 168N - (12N + 8) = 156N - 8, where N is the number of neurons included in the analysis. As the modulation function is not relevant for the L-model, its number of degrees of freedom is 168N - 10N = 158N.

Fitting procedure of the multisensory model

A maximum-likelihood method was used to fit our model to the data. The likelihood function—deviance (D: twice the negative log-likelihood ratio)—for one neuron is defined based on quasi-Poisson statistics[17,42,43] as follows.

$$D = \sum_{\hat{V}} \frac{2R_{obs}(\hat{V})}{\theta(\hat{V})}\left(\frac{R_{model}(\hat{V})}{R_{obs}(\hat{V})} - 1 - \log\frac{R_{model}(\hat{V})}{R_{obs}(\hat{V})}\right),$$

where $\hat{V}$ represents the four stimulus variables $\{\varphi_V, \varphi_A, I_V, I_A\}$, θ is a dispersion factor defined as $\theta(\hat{V}) = \frac{\sigma_{obs}{}^2(\hat{V})}{R_{obs}(\hat{V})}$, R_{obs} is the observed mean firing rate over trials, σ_{obs} is the SEM of the observed firing rate, and R_{model} is the estimated firing rate by the model described above. The sum is over all 168 non-redundant combinations of the stimulus patterns (see previous section). Deviance for the entire dataset is a sum of the individual deviances of all the modeled neurons.

Because of the global parameters that affect all the neurons, the fitting must be done simultaneously for all the neurons. In order to achieve the best accuracy of the modulation function, we included all the neurons that had either visual or auditory responses in the late time window. Since there are tens of responsive neurons per dataset, hundreds of parameters must be simultaneously fit to thousands of data points. We minimized the effort of this challenging fitting procedure by breaking down the function fit into smaller problems with a smaller number of parameters and repeating simple function fits. Namely, when the individual parameters of a neuron are changed, the global parameters are fixed, so the problem is reduced to 12 parameters for the neuron, and when the 8 global parameters are changed, the individual parameters for the neurons are fixed. These fittings are alternately repeated until the overall likelihood value converged (the exact condition is written below).

The fit was performed in two phases because doing the entire fit at once was unstable. For example, if both the neurons' effective RF azimuth (φ), and the modulation shapes are changed simultaneously, too many local minima are created and the fit cannot escape from them. By fixing the azimuths of the RF location in the first phase, we could stabilize the resulting properties of the modulation function (and its likelihood value was consistently better). In the second phase, some of the parameter boundaries were also relaxed (see Table 2).

The detailed fitting procedure is the following.

Phase 1: (narrow parameter boundaries)

1. Decide initial visual and auditory sigmoid parameters for individual neurons using unisensory response data.
2. Randomly set the modulation parameters in their allowable ranges (except for A_T that are set to a small value (0.1) to stabilize fitting).
3. Refit individual neurons using the full data and modulation.
4. Refit the modulation parameters using the updated individual cell parameters.
5. Repeat 3-4 until the change of NLL in one iteration is below the tolerance value (0.1).

6. Once it reaches the tolerance, perform a detailed error analysis (MINOS[44]) to further improve the fit and repeat 3-4 again.
7. Stop if the likelihood gain is less than the tolerant value after the MINOS analysis.

Phase 2: (relaxed parameter boundaries)

1. Perform the same procedure as 3-7 of Phase 1 with the relaxed parameter boundaries.

We repeated this fitting procedure 30 times with random initial modulation parameters, and used the fit that minimized the negative log-likelihood value. Once the fitting is completed, the error of each parameter was estimated from the Hessian matrix of this likelihood function. To calculate goodness-of-fit of individual neurons, the asymptotic equivalence between χ^2 and two times NLL (Wilks' theorem[45]) was assumed.

Figures

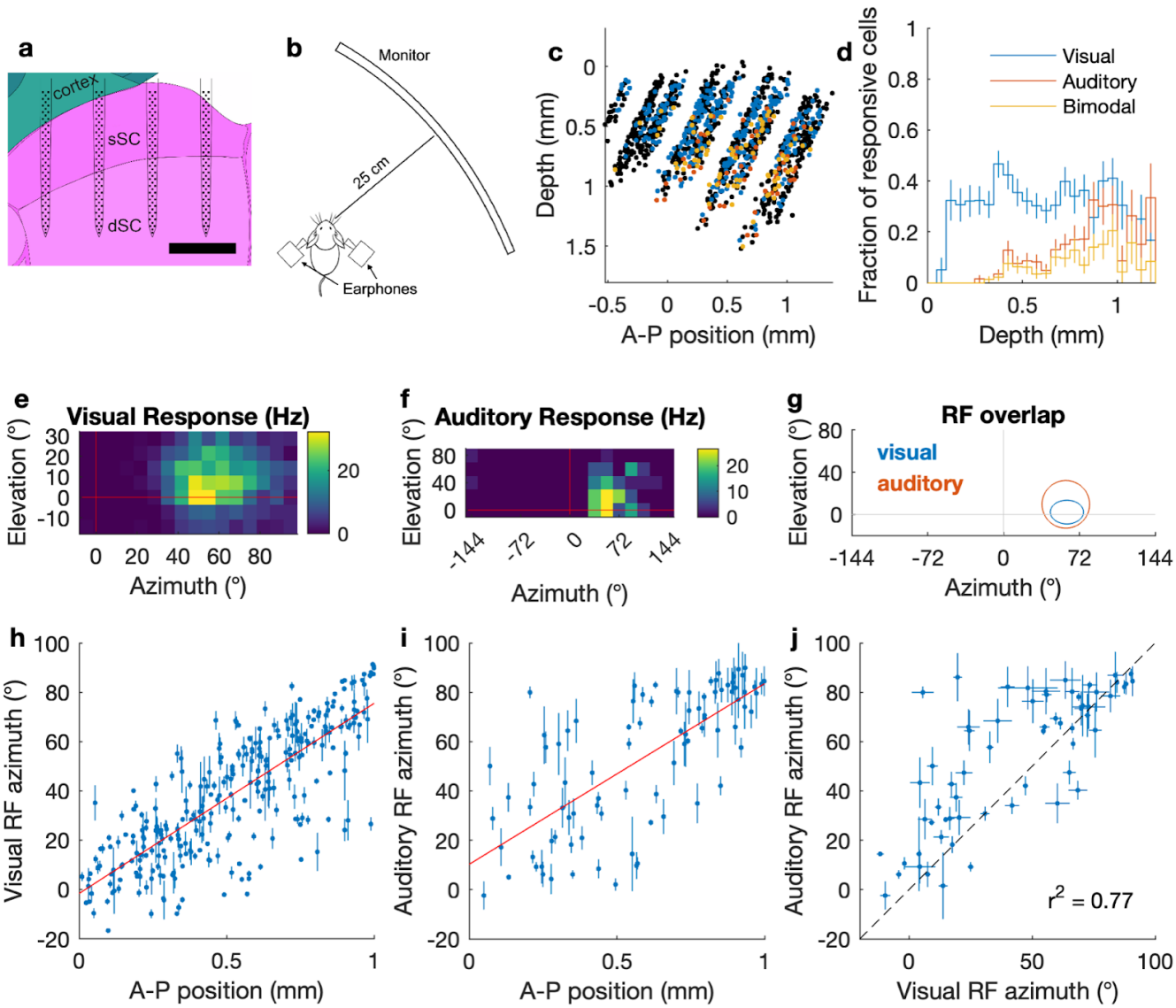


Figure 1: Topographic organization of visual and auditory responses in the mouse SC.

a: A schematic of a 256-electrode silicon probe in the superior colliculus (SC; sagittal plane). The background image is from Allen Mouse Reference Atlas[46]. sSC: superficial SC, dSC: deep SC. The scale bar is 500 μm.
b: A schematic of the recording setup. An awake head-fixed mouse is placed on a cylindrical treadmill. A computer display monitor is placed 25 cm away from the mouse. Virtual auditory space stimuli are delivered through earphones.
c: Reconstructed 2D map of all the cells with visual and/or auditory responses (blue: visual, red: auditory, yellow: bimodal, black: non-responsive; significant response with $p < 0.001$; quasi-Poisson statistics[17])
d: Fraction of the visual/auditory/bimodal neurons as a function of SC depth.
e-g: Spatial heat map of response firing rates of an example neuron that shows visual (e) and auditory (f) RFs, and a summary of their overlap (g). Ellipses in (g) indicate 1σ contours of 2D Gaussian fits.
h,i: Scatter plots of estimated RF azimuth vs. A-P SC position for visually responsive (h) and auditory responsive (i) SC neurons. The red line is a linear fit. The error bars represent 1-σ errors (68% CIs) inferred from function fitting.
j: A scatter plot of visual and auditory RF azimuths for bimodal neurons.

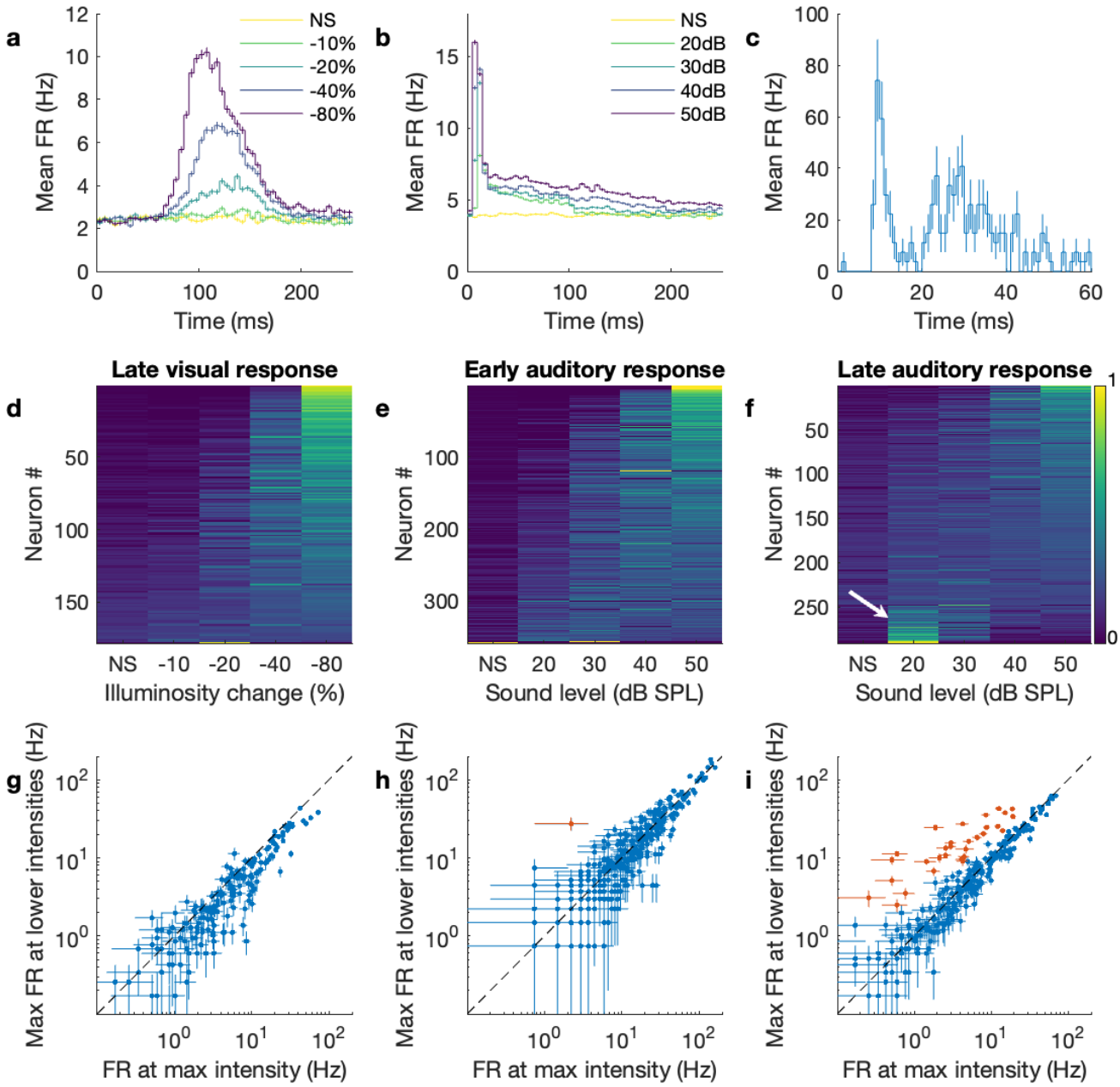


Figure 2. Temporal segregation and intensity profiles of SC visual and auditory responses
a, b: Average responses over time of all (a) visually or (b) auditory responsive neurons. Darker colors indicate higher contrast (visual stimuli) or stronger intensities (auditory stimuli; NS: no stimulus).
c: The auditory response of an example neuron over time (averaged over 30 trials at 85 grid locations) that has distinct early (<20ms) and late (>20ms) components. The error bars are SEM over the trials.
d: Normalized responses to different visual intensities. The neurons are sorted by the center of mass of the intensity response (d-f).
e: Normalized early responses to different auditory intensities.
f: Normalized late responses to different auditory intensities. Note that a fraction of cells prefer low intensities over high intensities (white arrow).

g-i: Scatter plots of the maximum (max) FR at non-max intensity vs the FR at max intensity (red points are non-monotonic neurons: $y > x$ with $p < 0.001$) of (g) the visual stimulus response; (h) the early auditory response; and (i) the late response to auditory stimuli. The error bars are SEM over 30 trials.

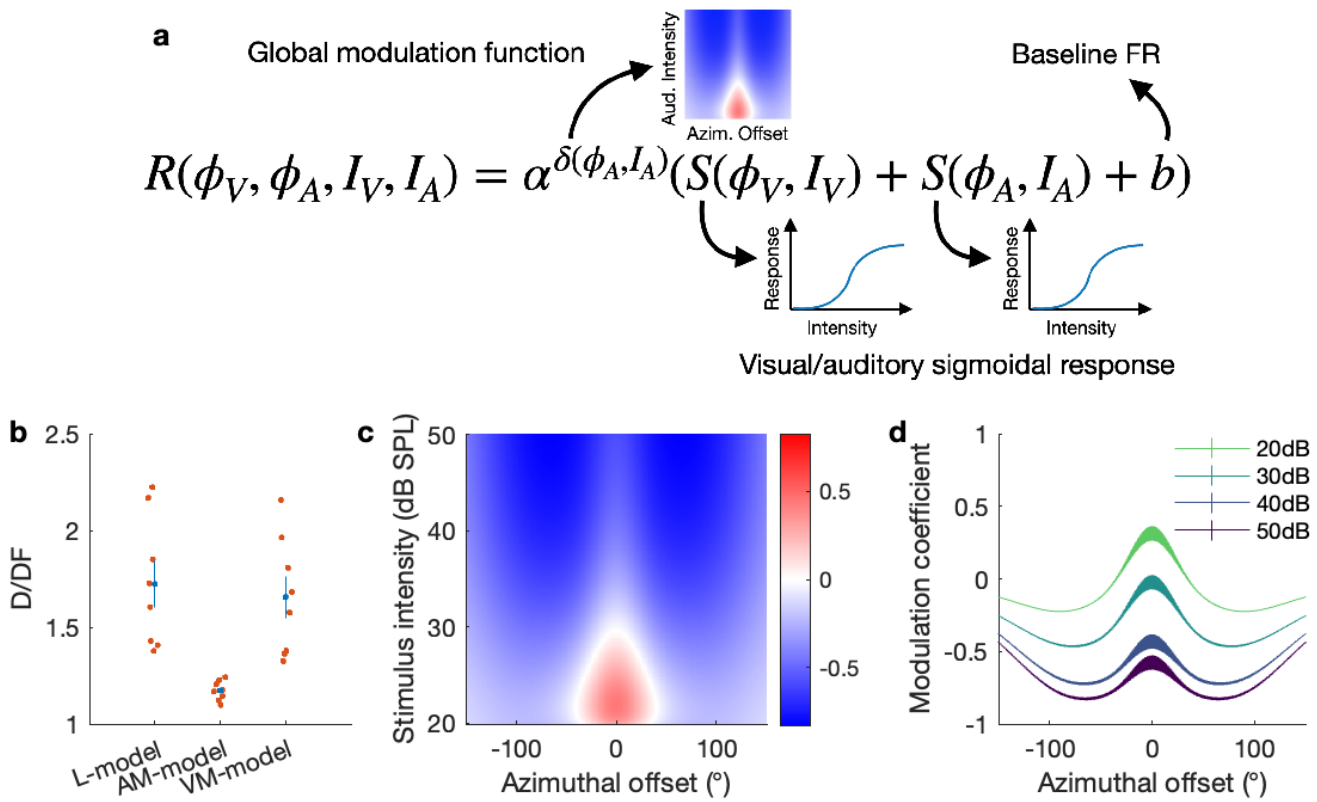


Figure 3: Function fit equation and results of modulation function

a: The equation of the model for A-dependent modulation. For the no-modulation model, Alpha is fixed to 1. For the V-dependent model, the delta depends on L_V and I_V.

b: The deviance per degree of freedom (D/DF) values of each model. Red dots are individual experiments. Blue is the mean and SEM.

c: Average modulation function from all the experiments. Superlinear addition could only occur in the area indicated by red, which is restricted to the small azimuthal offset range and the low auditory stimulus intensity.

d: Same as c for each dB value with error.

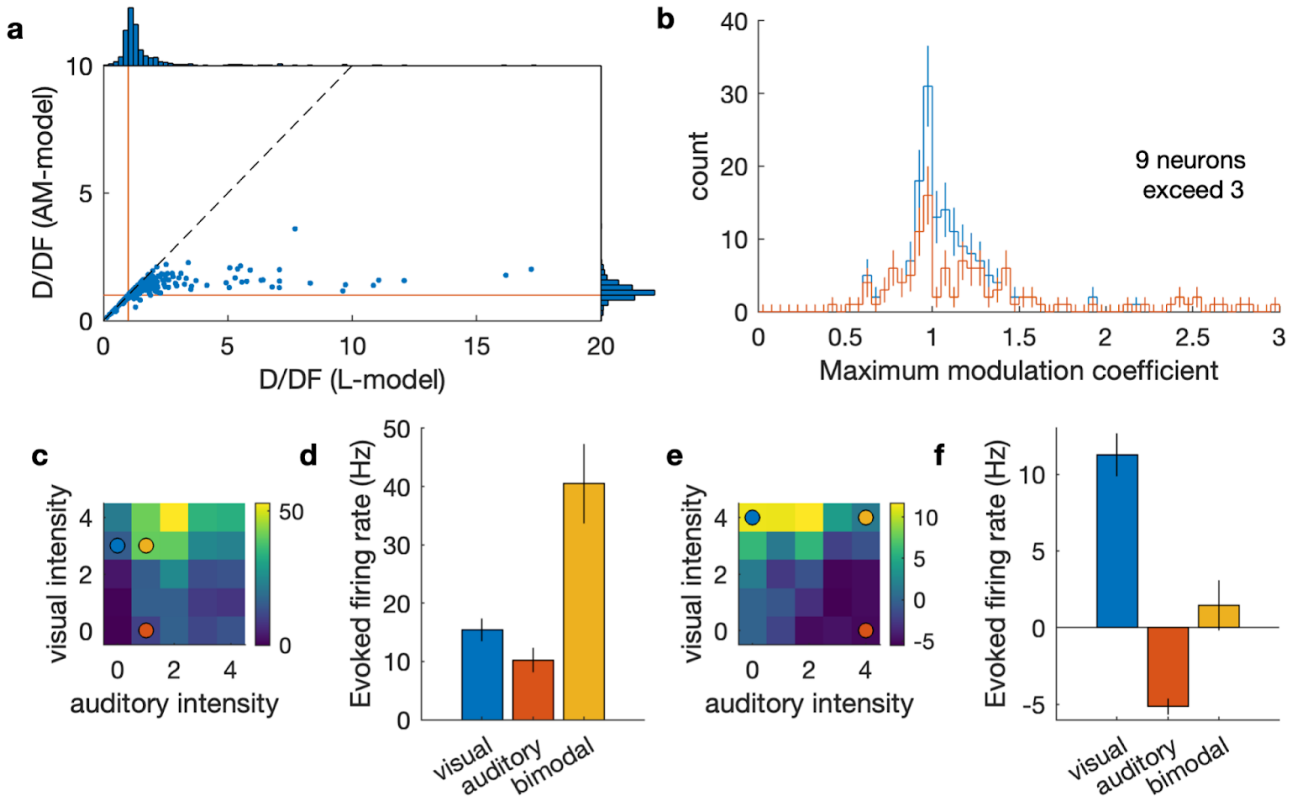


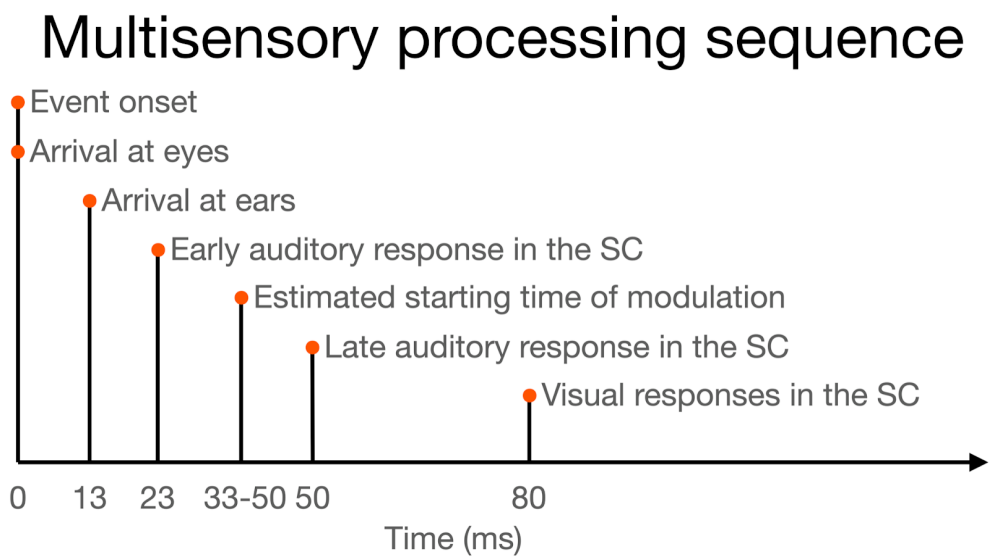
Figure 4: Visual/auditory interactions in the mouse SC

a: Deviance per degree of freedom (D/DF) of individual neurons. Red lines indicate D/DF = 1, good-fit lines. Many of the high D/DF neurons (bad-fit neurons) with the L-model have much smaller D/DF with the AM-model. There are few neurons that have high D/DF with the AM-model compared to the L-model—for example, there is only one neuron with D/DF > 3 in the AM-model, while there are 33 in the L-model—indicating that the AM-model is a good description of the data.

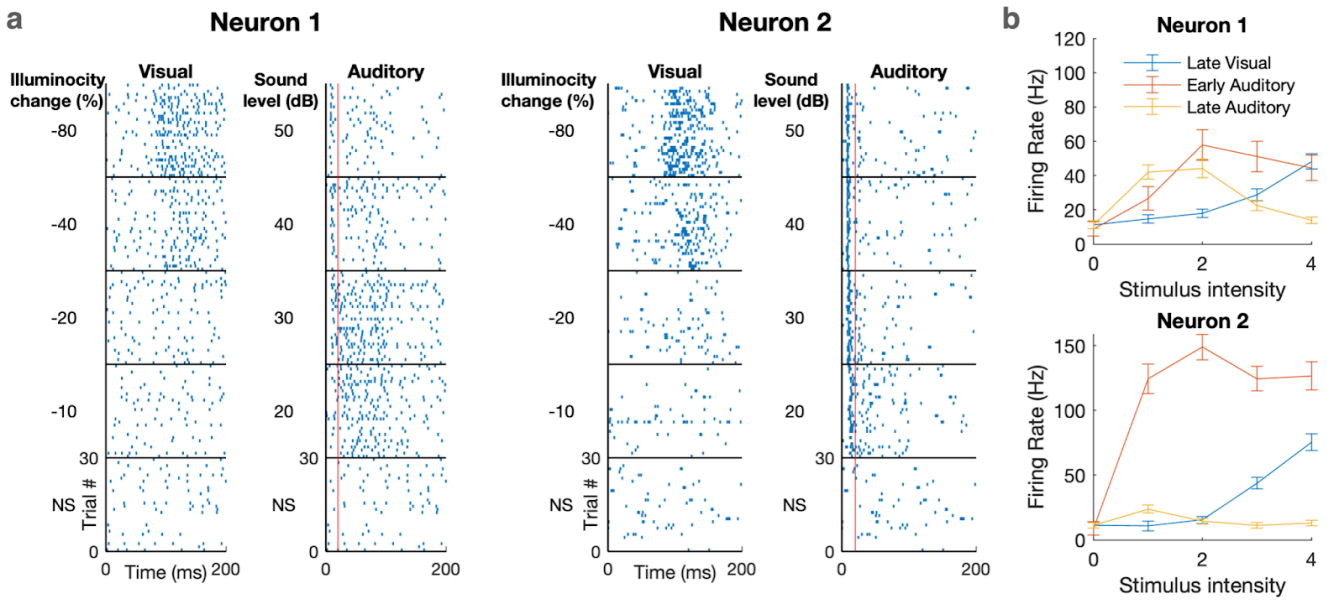
b: Histogram of the maximum modulation coefficient ($max(\alpha^{\delta})$) of 190 neurons with significant modulation ($\alpha > 1$ with p < 0.01). The red line is a histogram of neurons with the maximum modulation coefficient significantly (p < 0.01) different from 1. The modulation coefficient larger (less) than 1 indicates that the neuron exhibited superlinear (sublinear) addition of the visual and auditory responses.

c,d: Responses of an example neuron that exhibits superlinear multisensory integration viewed as a heat map (c) and with the firing rate of selected combinations of stimuli (colored dots) graphed in (d). A yellow circle in (c) indicates a combination of stimulus intensities with which significant superlinear enhancement is observed. Blue and red circles indicate corresponding unimodal stimuli. The details of the response of this neuron is elaborated in Supplementary Fig. 4.

e,f: Responses of an example neuron that exhibits multisensory suppression viewed as a heat map (e) and with the firing rate of selected combinations of stimuli (colored dots) graphed in (f). Colored circles indicate stimulus combinations with which multisensory suppression is observed.



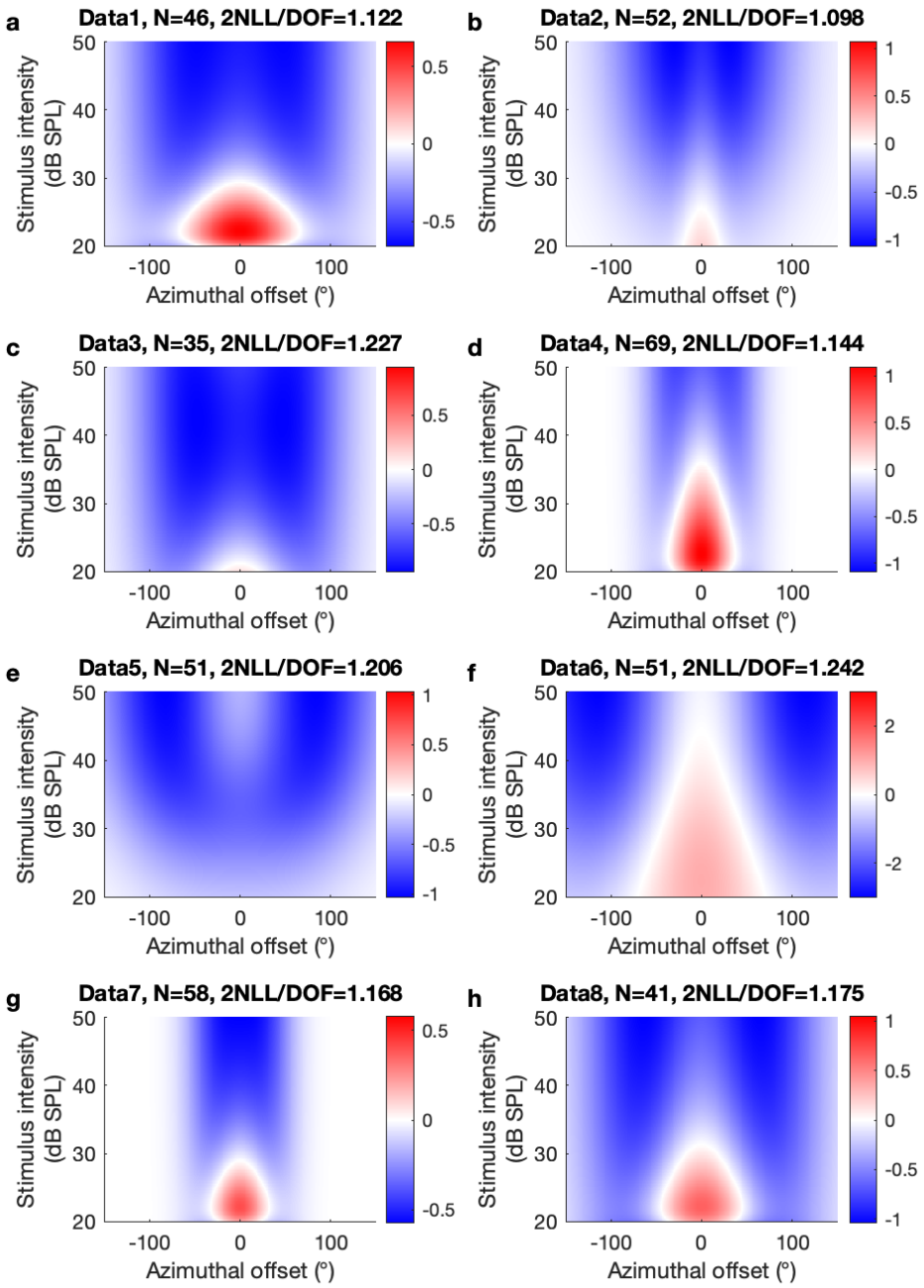
Supplementary Figure 1: Sequence and timing of events (stimuli and responses) in these multisensory experiments.



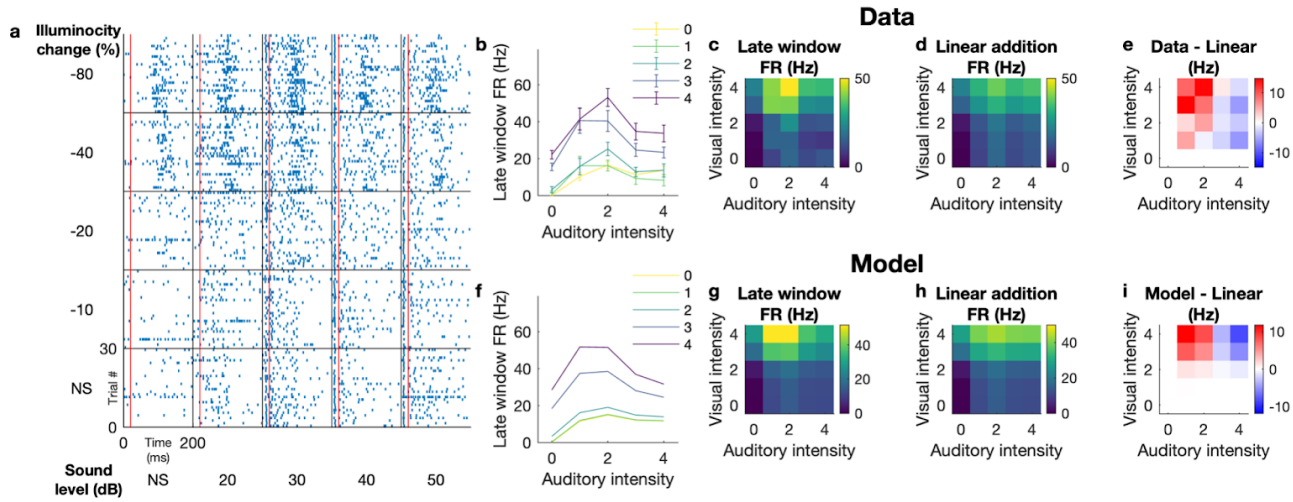
Supplementary Figure 2: Example responses of visuo-auditory bimodal neurons to changes in illumination and sound level.

a: Raster plots of two example neurons to different intensities of monomodal visual and auditory stimuli. These neurons exhibit non-monotonic auditory responses in the late time scale. The red line indicates 20 ms that separates the early and late time windows of auditory responses.

b: Summary firing rates of the two neurons in (a). The stimulus intensities are indicated by 0-4 for brevity; their actual values correspond with those used in (a). These neurons' peak early auditory response is not significantly larger than their max-intensity response ($p = 0.25$ for neuron 1; $p = 0.13$ for neuron 2), therefore, these are not considered 'non-monotonic' responses. On the other hand, the peak late-auditory response is significantly larger than the max-intensity response ($p = 4 \times 10^{-10}$ for neuron 1; $p = 0.004$ for neuron 2), therefore, these are considered 'non-monotonic' responses.



Supplementary Figure 3: Modulation functions of each individual dataset. Datasets a-d were used as the exploratory datasets while e-h were used as the blinded datasets (see methods).



Supplementary Figure 4: An example neuron that exhibits superlinear multisensory enhancement.
a: Raster plots to different combinations of visual and auditory stimuli. The red lines indicate 20 ms that separates the early and late time window for the auditory response.
b: Measured firing rates in the late time window. The colors indicate visual contrasts. The error bars are SEM over 30 trials. The stimulus intensities are indicated as 0-4 for brevity; these numbers correspond with the values in (a).
c: A heat map representation of the late window firing rate.
d: A heat map of the linear additions of the firing rates to unimodal visual stimuli (left most column) and auditory stimuli (bottom row).
e: Difference of the firing rates in response to multisensory stimuli (c) from linear addition of the unimodal firing rates (d). This neuron exhibits superlinear addition when the auditory stimulus is weak and sublinear addition when the auditory stimulus is strong.
f-i: Modeled responses for this neuron. The model captures complex non-linear responses to different intensity levels, as well as a shift from superlinear regime to sublinear regime as a function of the auditory intensity.

References

1. Stein, B. E. & Meredith, M. A. The merging of the senses. *Cogn. Neurosci.* **211**, (1993).
2. May, P. J. The mammalian superior colliculus: laminar structure and connections. *Prog. Brain Res.* **151**, 321–378 (2006).
3. King, A. J., Schnupp, J. W. H. & Thompson, I. D. Signals from the Superficial Layers of the Superior Colliculus Enable the Development of the Auditory Space Map in the Deeper Layers. *J. Neurosci.* **18**, 9394–9408 (1998).
4. Knudsen, E. I. Auditory and visual maps of space in the optic tectum of the owl. *J. Neurosci.* **2**, 1177–1194 (1982).
5. Brainard, M. S. & Knudsen, E. I. Sensitive Periods for Visual Calibration of the Auditory Space Map in the Barn Owl Optic Tectum. *J. Neurosci.* **18**, 3929–3942 (1998).
6. Knudsen, E. I. Control from below: the role of a midbrain network in spatial attention. *Eur. J. Neurosci.* **33**, 1961–1972 (2011).
7. Knudsen, E. I. Midbrain and forebrain systems for bottom-up control of spatial attention. in *The neuroscience of attention: Attentional control and selection, (pp* (ed. Mangun, G. R.) vol. 278 131–150 (Oxford University Press, xiii, 2012).
8. Meredith, M. A. & Stein, B. E. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology* vol. 56 640–662 (1986).
9. Meredith, M. A. & Stein, B. E. Interactions among converging sensory inputs in the superior colliculus. *Science* **221**, 389–391 (1983).
10. Alex Meredith, M. & Stein, B. E. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res.* **365**, 350–354 (1986).
11. Meredith, M. A., Nemitz, J. W. & Stein, B. E. Determinants of multisensory integration in superior

colliculus neurons. I. Temporal factors. *J. Neurosci.* **7**, 3215–3229 (1987).

12. Wallace, M. T., Wilkinson, L. K. & Stein, B. E. Representation and integration of multiple sensory inputs in primate superior colliculus. *J. Neurophysiol.* **76**, 1246–1266 (1996).
13. Zahar, Y., Reches, A. & Gutfreund, Y. Multisensory enhancement in the optic tectum of the barn owl: spike count and spike timing. *J. Neurophysiol.* **101**, 2380–2394 (2009).
14. Populin, L. C. & Yin, T. C. T. Bimodal interactions in the superior colliculus of the behaving cat. *J. Neurosci.* **22**, 2826–2834 (2002).
15. Wang, L., Sarnaik, R., Rangarajan, K., Liu, X. & Cang, J. Visual receptive field properties of neurons in the superficial superior colliculus of the mouse. *J. Neurosci.* **30**, 16573–16584 (2010).
16. Ito, S., Feldheim, D. A. & Litke, A. M. Segregation of Visual Response Properties in the Mouse Superior Colliculus and Their Modulation during Locomotion. *J. Neurosci.* **37**, 8428–8443 (2017).
17. Ito, S., Si, Y., Feldheim, D. A. & Litke, A. M. Spectral cues are necessary to encode azimuthal auditory space in the mouse superior colliculus. *Nat. Commun.* **11**, 1087 (2020).
18. Palmer, A. R. & King, A. J. The representation of auditory space in the mammalian superior colliculus. *Nature* **299**, 248–249 (1982).
19. Wallace, M. T. & Stein, B. E. Development of Multisensory Neurons and Multisensory Integration in Cat Superior Colliculus. *J. Neurosci.* **17**, 2429–2444 (1997).
20. King, A. J. & Hutchings, M. E. Spatial response properties of acoustically responsive neurons in the superior colliculus of the ferret: a map of auditory space. *J. Neurophysiol.* **57**, 596–624 (1987).
21. Kasai, M. & Isa, T. Imaging population dynamics of surround suppression in the superior colliculus. *Eur. J. Neurosci.* **44**, 2543–2556 (2016).
22. Mysore, S. P. & Kothari, N. B. Mechanisms of competitive selection: A canonical neural circuit framework. *Elife* **9**, e51473 (2020).
23. Russell, J. T. Imaging calcium signals in vivo: a powerful tool in physiology and pharmacology. *Br. J. Pharmacol.* **163**, 1605–1625 (2011).

24. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).

25. Shobe, J. L., Claar, L. D., Parhami, S., Bakhurin, K. I. & Masmanidis, S. C. Brain activity mapping at multiple scales with silicon microprobes containing 1,024 electrodes. *Journal of Neurophysiology* vol. 114 2043–2052 (2015).

26. Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., Jr & Rowland, B. A. Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Exp. Brain Res.* **198**, 113–126 (2009).

27. Mazurek, M., Kager, M. & Van Hooser, S. D. Robust quantification of orientation selectivity and direction selectivity. *Front. Neural Circuits* **8**, 92 (2014).

28. Phongphanphanee, P. *et al.* Distinct local circuit properties of the superficial and intermediate layers of the rodent superior colliculus. *Eur. J. Neurosci.* **40**, 2329–2343 (2014).

29. Chabot, N., Mellott, J. G., Hall, A. J., Tichenoff, E. L. & Lomber, S. G. Cerebral origins of the auditory projection to the superior colliculus of the cat. *Hear. Res.* **300**, 33–45 (2013).

30. Tokuoka, K., Kasai, M., Kobayashi, K. & Isa, T. Anatomical and electrophysiological analysis of cholinergic inputs from the parabigeminal nucleus to the superficial superior colliculus. *J. Neurophysiol.* **124**, 1968–1985 (2020).

31. Deichler, A. *et al.* A specialized reciprocal connectivity suggests a link between the mechanisms by which the superior colliculus and parabigeminal nucleus produce defensive behaviors in rodents. *Sci. Rep.* **10**, 16220 (2020).

32. Wang, Y., Major, D. E. & Karten, H. J. Morphology and connections of nucleus isthmi pars magnocellularis in chicks (Gallus gallus). *J. Comp. Neurol.* **469**, 275–297 (2004).

33. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).

34. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).

35. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
36. Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? (2007).
37. Gale, S. D. & Murphy, G. J. Active Dendritic Properties and Local Inhibitory Input Enable Selectivity for Object Motion in Mouse Superior Colliculus Neurons. *J. Neurosci.* **36**, 9111–9123 (2016).
38. MacCoun, R. & Perlmutter, S. Blind analysis: Hide results to seek the truth. *Nature* **526**, 187–189 (2015).
39. Klein, J. R. & Roodman, A. BLIND ANALYSIS IN NUCLEAR AND PARTICLE PHYSICS. *Annu. Rev. Nucl. Part. Sci.* **55**, 141–163 (2005).
40. Litke, A. M. *et al.* What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity. *IEEE Trans. Nucl. Sci.* **51**, 1434–1440 (2004).
41. Wiltschko, A. B., Gage, G. J. & Berke, J. D. Wavelet filtering before spike detection preserves waveform shape and enhances single-unit discrimination. *J. Neurosci. Methods* **173**, 34–40 (2008).
42. Ver Hoef, J. M. & Boveng, P. L. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* **88**, 2766–2772 (2007).
43. Wedderburn, R. W. M. Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method. *Biometrika* **61**, 439–447 (1974).
44. James, F. & Roos, M. Minuit - a system for function minimization and analysis of the parameter errors and correlations. *Comput. Phys. Commun.* **10**, 343–367 (1975).
45. Wilks, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Ann. Math. Stat.* **9**, 60–62 (1938).
46. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

Acknowledgements

This work was supported by Brain Research Seed Funding provided by UCSC, the National Institutes of Health (2R01EY022117, R21 EY032230-01), and a donation from John Chen to A.M.L.. We thank Sotiris Masmanidis for providing us with the silicon probes; and Jena Yamada for helpful comments on the manuscript.

Author Contributions

S.I., D.A.F. and A.M.L. conceived the project. S.I. designed the experiments. S.I. and Y.S. acquired the data. S.I. and Y.S. analyzed the data. S.I., D.A.F. and A.M.L interpreted the results and wrote the manuscript.

These authors jointly supervised this work: David A. Feldheim, Alan M. Litke

Competing interests

The authors declare no competing interests.