Functional specialisation of multisensory temporal integration in the mouse superior colliculus

Bianchini Gaia[1], Ines Razafindrahaba[1], Marcelo J. Moglie[1], George Konstantinou[1], Xavier Cano-Ferrer[1], Albane Imbert[1], M. Florencia Iacaruso[1]

[1]*The Francis Crick Institute, London, UK.*

Our perception of the world depends on the brain's ability to integrate information from multiple senses, with temporal disparities providing a critical cue for binding or segregating cross-modal signals[1,2]. The superior colliculus (SC) is a key site for integrating sensory modalities, but how cellular and network mechanisms in distinct anatomical regions within the SC contribute to multisensory integration remains poorly understood. Here, we recorded responses from over 5,000 neurons across the SC's anatomical axes of awake mice during presentations of spatially coincident audiovisual stimuli with varying temporal asynchronies. Our findings revealed that multisensory neurons reliably encoded audiovisual delays and exhibited nonlinear summation of auditory and visual inputs, with nonlinearities being more pronounced when visual stimuli preceded auditory stimuli, consistent with the natural statistics of light and sound propagation. Nonlinear summation was crucial for population-level decoding accuracy and precision of AV delay representation. Moreover, enhanced population decoding of audiovisual delays in the posterior-medial SC, facilitated temporal discriminability in the peripheral visual field. Cross-correlation analysis indicated higher connectivity in the medial SC and functional specific recurrent connectivity, with visual, auditory, and multisensory neurons preferentially connecting to other neurons of the same functional subclass, and multisensory neurons receiving approximately 50 percent of the total local input from other multisensory neurons. Our results highlight the interplay between single-neuron computations, network connectivity, and population coding in the SC, where nonlinear integration, distributed representations and regional functional specialisations enables robust sensory binding and supports the accurate encoding of temporal multisensory information. Our study provides new insights into how the brain leverages both single-neuron and network-level mechanisms to represent sensory features by adapting to the statistics of the natural world.

Introduction

The evolution of neural circuits has shaped their structure and ability to extract information from the physical world, optimising organisms' survivability and reproductive success. To exploit environmental statistical regularities, the brain integrates information from various sensory modalities, enhancing the precision, accuracy, and processing speed of perception[3–8]. Temporal delays between information from the different sensory modalities provide a crucial cue for determining which cross-modal signals belong together[1,2]. However, the relative timing of signals from an external event can vary significantly between sensory modalities. The difference in velocities of light and sound introduces a lag for auditory signals relative to visual information. Since the speed of light and sound through air are constant, their time difference can provide absolute information about the source distance, independent of the environment. Indeed, in humans, audiovisual (AV) temporal disparity can influence distance estimation[9,10]. In addition to temporal disparities arising from physical variables in the external world, the temporal characteristics of stimuli that activate peripheral receptor cells differ considerably between sensory systems. As a result, the timescales of signals can differ by orders of magnitude across sensory systems. Thus, multisensory integration requires the comparison and linking of information across different modalities and timescales[11].

The superior colliculus (SC) receives sensory information from multiple modalities and is critical for reorienting an organism towards objects of interest and controlling innate behaviours[12–17]. Multisensory neurons in the SC respond to combinations of visual, auditory, and/or somatosensory stimuli, with their receptive fields arranged to form topographically aligned spatial maps[18–20]. Interestingly, visual-auditory neurons in the SC show nonlinear integration of cross-modal information[21,22] and enhanced responses at variable delays between visual and auditory stimuli[23,24]. Such temporal integration may vary along the SC's topographic map, reflecting functional specialisations for distinct behavioural responses. Rodent studies have revealed that gene expression varies along the antero-posterior axis of the SC[12] and show functional and anatomical segregation of inputs and outputs[12,25–28]. The medial SC encodes information about the upper visual field and promotes defensive responses, while the lateral region encodes information about the lower visual field and evokes approach and appetitive behavior[25–28]. Whether these anatomically distinct regions of the SC exhibit specialisation in multisensory temporal integration to facilitate different behavioural responses remains unexplored.

To investigate how the SC represents AV temporal disparities, we recorded activity from approximately 5,000 neurons across the SC of awake mice while presenting spatially coincident AV stimuli with varying temporal delays. We found that individual neurons exhibit a broad tuning to AV delays and nonlinear summation of unisensory inputs. At the population level, decoding accuracy for AV delays was highest in the postero-medial SC, which encodes the peripheral visual field. This enhanced temporal discriminability relies on the nonlinear integration of multisensory inputs in single neurons. Cross-correlation analysis revealed that multisensory neurons might receive a significant proportion of local input from other multisensory neurons, suggesting that recurrent local connections might contribute to nonlinear integration of cross-modal inputs across different timescales. Our results suggest that region-specific functional specialisation in the SC underlies multisensory temporal integration, enhancing the encoding of AV information to guide appropriate orienting responses.

Results

Spatially aligned visual and auditory receptive fields in the mouse superior colliculus

To study the multisensory representation of space in the mouse SC, we recorded the activity of neuronal populations at a range of tissue depths and anatomical locations by performing extracellular electrophysiological recordings using Neuropixels 1.0 probes[29] in head-fixed awake mice (**Figure 1A**). Recognising that sensory processing varies across different regions of the SC[30,31] we performed a comprehensive characterisation of AV integration by recording from different sites along the antero-posterior (AP), medio-lateral (ML) and dorso-ventral (DV) axes. We covered 1.7 mm, 1.9 mm and 2.3 mm along the AP, ML, and DV axes of the SC, respectively. Anatomical locations of recording sites were confirmed by combining electrophysiological features with histological reconstructions of fluorescently-labelled probe tracks. We used a custom-made hemispheric stimulation apparatus to

provide spatially coincident visual and auditory stimuli from a range of spatial locations (**Figure 1B**). The stimuli consisted of a 100 ms white noise (1–22 kHz) and a 100 ms light flash presented via a dimmable light-emitting diodes (LED) positioned at the centre of each speaker. Sound level was fixed at 60dB to enhance multisensory interactions at lower stimulus intensities[21] and to minimise potential confounds from sound-evoked body motion affecting auditory responses **(Suppl. Figure 1)** [32,33]. We found that 13% of the recorded neurons responded to visual stimuli, 9% to auditory stimuli and 34% were classified as multisensory (**Figure 1C**). Multisensory neurons exhibited a variety of response patterns, including neurons responsive to both modalities when presented independently (bimodal, 10%) and neurons responsive only to multisensory stimulation (gated, 16%). These neuronal subclasses were differentially distributed along the depth of the SC ($P = 1 \times 10^{-145}$, Kruskal-Wallis; **Figure 1D**). Consistent with previous reports[22,34], visual neurons were primarily located in the superficial layers (mean depth 756 ± 318 µm), while auditory neurons were found in the deeper layers of the SC (1388 ± 401 µm). Bimodal neurons were located more superficially (992 ± 458 µm), whereas gated neurons were predominantly situated in the deeper layers of the SC (1330 ± 535 µm).

To verify that our stimulation apparatus could capture the topographic alignment of visual and auditory maps in the SC[22,34] , in a subset of experiments, we estimated the location of the centre of the receptive field (RF) of individual neurons by stimulating at 7 azimuths and 5 elevations with 18° steps surrounding the centre of the population RF, determined online. We observed that the RF centroid of visual and auditory responsive neurons recorded simultaneously with a vertical insertion of the Neuropixels probe were spatially aligned (**Figure 1E,F**). As a result, the receptive fields obtained from the averaged population responses to visual and auditory stimuli of simultaneously recorded neurons were also aligned (**Figure 1G,H,** N = 11 recordings, Pearson's correlation coefficient Azimuth: R = 0.86, $P = 1 \times 10^{-4}$; Elevation: R = 0.76, P = 0.1). Moreover, bimodal neurons exhibited overlapping unisensory visual and auditory receptive fields in azimuth (**Figure 1I,J**; n = 89 neurons, Azimuth: R= 0.4, $P = 1 \times 10^{-4}$; Elevation: R = 0, P = 0.9). Additionally, for these neurons the RFs determined from multisensory stimulation were spatially aligned to the RF estimated from the average responses to the unisensory stimuli (**Figure 1K,** Azimuth: R = 0.56, $P = 1 \times 10^{-9}$; Elevation: R = 0.24, P = 0.02), and the average population RF estimated from unisensory responses matched the multisensory RF (**Figure 1K,L,** Azimuth: R = 0.87, $P = 1 \times 10^{-4}$; Elevation: R = 0.69, P = 0.01). Overall, these results confirm the spatial alignment of visual, auditory and audiovisual RFs in the mouse SC, mapped through unisensory and multisensory stimulation.

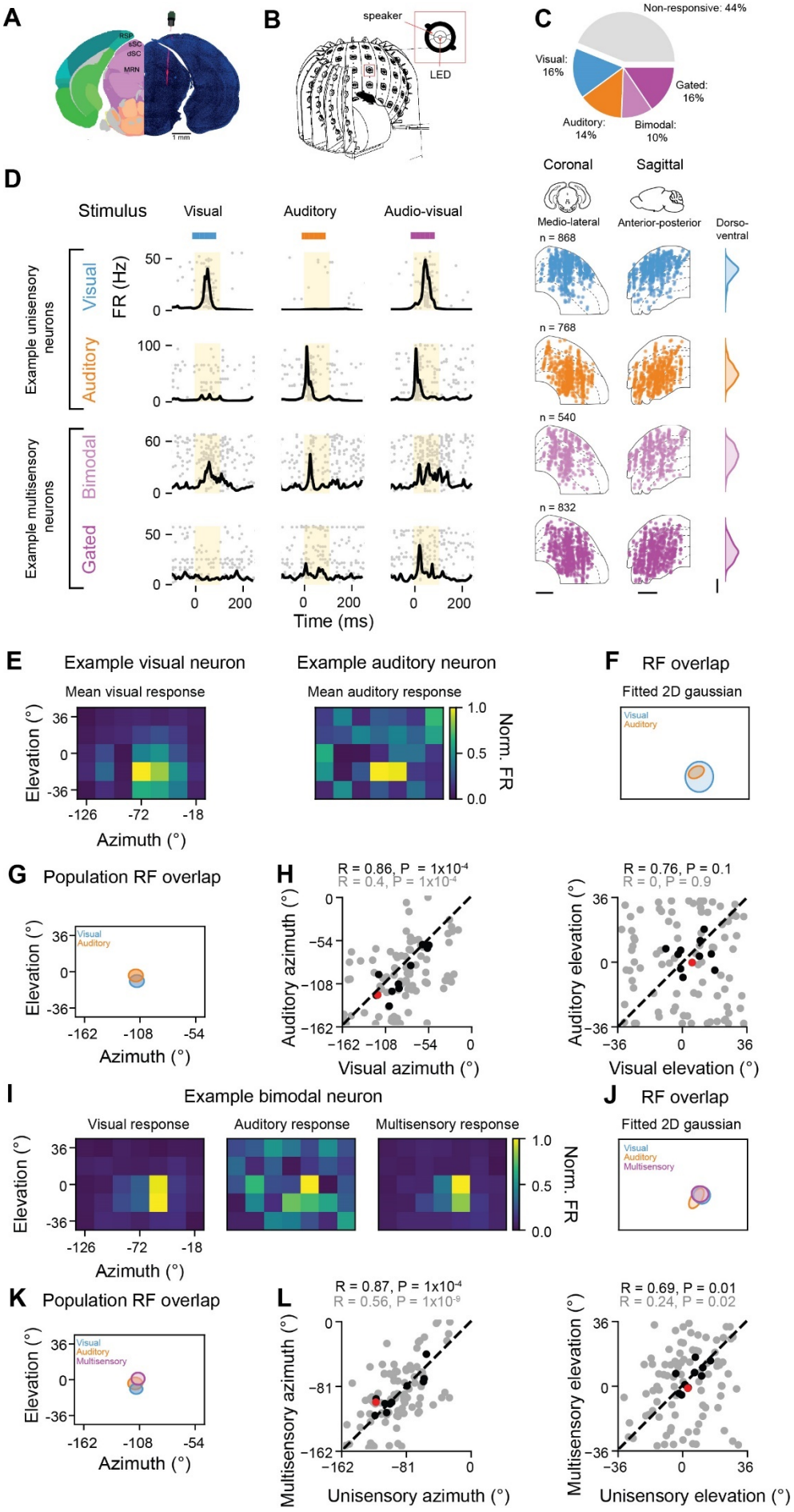

A
RSP
sSC
dSC
MRN
1 mm
B
speaker
LED
C
Non-responsive: 44%
Visual: 16%
Gated: 16%
Auditory: 14%
Bimodal: 10%
D
Stimulus
Visual
Auditory
Audio-visual
Example unisensory neurons
Example multisensory neurons
Visual
Auditory
Bimodal
Gated
FR (Hz)
Time (ms)
Coronal
Sagittal
Medio-lateral
Anterior-posterior
Dorso-ventral
n = 868
n = 768
n = 540
n = 832
E
Example visual neuron
Example auditory neuron
Mean visual response
Mean auditory response
Elevation (°)
Azimuth (°)
Norm. FR
F
RF overlap
Fitted 2D gaussian
G
Population RF overlap
H
R = 0.86, P = 1x10-4
R = 0.4, P = 1x10-4
Auditory azimuth (°)
Visual azimuth (°)
R = 0.76, P = 0.1
R = 0, P = 0.9
Auditory elevation (°)
Visual elevation (°)
I
Example bimodal neuron
Visual response
Auditory response
Multisensory response
J
RF overlap
Fitted 2D gaussian
K
Population RF overlap
L
R = 0.87, P = 1x10-4
R = 0.56, P = 1x10-9
Multisensory azimuth (°)
Unisensory azimuth (°)
R = 0.69, P = 0.01
R = 0.24, P = 0.02
Multisensory elevation (°)
Unisensory elevation (°)

Figure 1 | Spatially aligned visual and auditory receptive fields in the mouse superior colliculus. A. Histological reconstruction of the probe track with 4′,6-diamidino-2-phenylindole (DAPI; blue) and DiI (red) staining for an example recording site and corresponding alignment to the Allen Mouse Brain Atlas. RSP: retrosplenial cortex; sSC: superficial layers of the superior colliculus; dSC: deep layers of the superior colliculus; MRM: midbrain reticular nucleus. **B**. Experimental setup. Light flashes and ~60-65 dB white noise were presented to awake head-fixed mice using a custom hemispheric stimulation apparatus to provide spatially coincident visual and auditory stimuli. **C**. Proportion of neurons recorded in the SC, subdivided into non-responsive, visual, auditory, bimodal and gated. **D**. Left: Response profiles for example unisensory and multisensory neurons. The spike raster plots (spiking activity for 50 repetitions) are overlaid with the peristimulus time histogram (PSTH) during visual (blue bar), auditory (orange bar) and simultaneous audiovisual (AV) stimulation (magenta bar). Right: Anatomical location of all recorded neurons from each subclass along the medio-lateral (ML, left) and antero-posterior (AP, centre) axes of the SC. Distribution along the dorso-ventral (DV) axis is shown as a darnel density estimate plot (right). Scale bar: 0.5 mm. **E**. Spatial heat map of normalised firing rates of two simultaneously recorded visual (left) and auditory (right) unisensory neurons. **F**. Overlap between visual (blue) and auditory (orange) RFs of neurons shown in **E**. Ellipses indicate contours of 2D Gaussian fits. **G**. Averaged population visual (blue) and auditory (orange) RFs spatial coordinates obtained from an example recording. Size of the ellipses denote 90% confidence intervals around the mean RF coordinates. **H**. Correlation between visual and auditory RF azimuth (left) and elevation (right) for individual neurons (gray circles) and for individual recordings (black circles). Red circle corresponds to the example recording shown in **G**. N = 9 recordings, P values from Pearson Correlation coefficient. **I**. same as **E** for an example bimodal neuron for visual (left), auditory (centre) and multisensory (right) stimulation. **J**. Overlap between visual, auditory and multisensory (magenta) RFs of neuron in **I**. **K**. Overlap between visual (blue), auditory (orange) and multisensory (magenta) averaged population RFs spatial coordinates obtained from an example recording. Size of the ellipses indicate 90% confidence intervals around the mean RF coordinates. **L**. Correlation between multisensory and mean unisensory azimuth (preferred unisensory modality determined independently for each neuron, left) and elevation (right) for individual neurons (gray circles) and for the population (black circles). Red dot corresponds to the example recording shown in **K**.

Temporal disparities between audio and visual inputs are encoded in superior colliculus neurons

To investigate how intermingled SC populations integrate audiovisual information across different timescales, we presented auditory and visual stimuli at congruent spatial locations while systematically varying their temporal relationship. To determine the range of temporal disparities to test, we first studied the latency of the neuronal responses to unisensory visual and auditory stimuli. While we found maximum responses in visual neurons evoked by either the start (onset) or the end (offset) of the visual stimulus, auditory neurons primarily responded to the onset of the auditory stimulus (**Figure 2A**). The difference in response onset between visual and auditory neurons was approximately 50 ms (**Figure 2A,** median visual onset latency = 69 ms, interquartile range (IQR) = 35 ms; median auditory onset latency = 21 ms, IQR = 22 ms, $P = 1 \times 10^{-26}$, Kruskal-Wallis). This difference in response latency could result in a maximal multisensory integration when the visual stimulus precedes the auditory by this amount of time. We therefore presented unisensory visual and auditory stimuli, with varying AV delays ranging from 0 and 100 ms, ensuring that the visual stimulus always preceded the auditory one (**Figure 2B**). Importantly, sound evoked uninstructed body movements were similar across all multisensory trial types (P = 0.7, Kruskal-Wallis, **Suppl. Figure 1**).

Overall, we recorded a total of 5360 neurons from 24 mice (92 recording sites) and assessed their responsiveness using a ZETA test[35]. We found that 25% of the multisensory neurons (corresponding to 12% of the total recorded population) were modulated by AV delay (see Methods). Within the multisensory population, AV delay-selective neurons constituted 23% of the bimodal group and 26% of the gated group. For all neurons that showed statistically significant differences between their multisensory and unisensory responses, we determined their preferred delay as the condition that evoked the highest peak firing rate (see Methods). Overall, AV delay-selective neurons preferring different delays were intermingled with other neuronal groups and there was no particular bias in their distribution along the different anatomical axes of the SC (**Figure 2C,** mean depth 382 ± 143 µm, Spatial Autoregressive Model, SAM: P = 0.1; mean Moran Index = 0). **Figure 2D,E** shows the peristimulus time histograms (PSTHs) and corresponding tuning curves for two example neurons. Consistent with previous reports[23,24], across the population of SC neurons the mean preferred AV delay was 54 ± 30

ms (**Figure 2F**). However, the distribution was broad, and we found neurons preferring each of the tested AV delays.

To characterise the selectivity of neuronal responses to specific AV delays, we calculated a specificity index (**Figure 2G**, see methods) and found that nearly half of the delay-selective neurons (48%) exhibited a high specificity index (> 0.9). The degree of firing rate (FR) modulation by AV delay, was characterised using a FR modulation index (**Figure 2H**, see methods). Most neurons showed a modest increase in firing rate at the preferred delay condition compared to the delay condition evoking the weakest response, with only 9% of neurons exhibiting an index equal to or exceeding 0.5 (median = 0.28, IQR = 0.17). Furthermore, the reliability of responses at the preferred delay was relatively low, with only 7% of neurons showing a high trial-to-trial correlation coefficient (> 0.3, median = 0.10, IQR = 0.09; **Figure 2I**). Nonetheless, the reliability of individual neurons across trials was higher than chance for 99% of the delay-selective neurons (chance median = -0.0002, IQR = 0.003, see methods). This demonstrates that the multisensory responses were reliable and specific to audiovisual delays. This reliability and specificity were comparable across the different AV delays tested (selectivity index: P = 0.14; FR modulation index: P = 0.66; reliability index: P = 0.67, Kruskal-Wallis). Together, these results show that a wide range of AV delays are represented in the responses of individual SC neurons, with a high degree of specificity.

Given the similarity of the mean preferred AV delay across the population and the difference in the mean auditory and visual latencies, we sought to determine whether AV delay selectivity could be predicted from responses to unisensory inputs at single neuron level. However, as previously reported[24], the preferred AV delay of individual neurons was not simply predicted by the difference in latency between their responses to unisensory stimuli (**Figure 2J,** P = 0.7; least squares regression), nor by their latency of the unisensory responses alone (**Suppl. Figure 2,** visual latency: P = 0.9; auditory latency: P = 0.6). To assess if the temporal dynamics of the unisensory response profiles are indicative of AV delay selectivity, we linearly added the firing rates of the visual and the auditory response, after shifting the latter by the different delays experimentally tested, and estimated the peak firing rate of these summed AV responses. The AV delay with maximum peak firing rate from the linearly summed responses was selected as the predicted preferred delay of the neuron (see Methods, **Figure 2K**). We found a weak correlation between the observed and predicted preferred delay (Pearson's correlation coefficient = 0.05, P = 0.08). Overall, the preferred delay was higher than the predicted AV delay (mean ± s.d., observed: 54 ± 30 ms; predicted: 41 ± 32 ms; **Figure 2L**). Interestingly, discrepancies between the predicted and observed responses revealed nonlinearities in the summation of auditory and visual inputs. To characterise these nonlinearities across AV delays, we calculated the multisensory interaction index (MII, see Methods). Nonlinear interactions were classified as supralinear when the multisensory response exceeded the sum of unisensory responses and sublinear when it fell short. Specifically, neurons that exhibited a 50% increase in firing rate relative to the sum of unisensory responses (MII > 0.5) were classified as exhibiting supralinear interactions, while those with an 50% reduction (MII < -0.5) were classified as sublinear.

Linear summation characterised the response of most neurons when auditory and visual stimuli were presented simultaneously (AV_0, median MII = 0.00, IQR = 0.33, **Figure 2M**), while a small proportion of neurons displayed supralinear (8.9%) or sublinear (1.2%) multisensory responses (**Figure 2N**). Notably, longer delays resulted in larger median MII across the population of recorded neurons ($P = 1 \times 10^{-15}$, Kruskal-Wallis). When considering only the preferred delay for each neuron, 33% of neurons exhibited supralinear summation of unisensory inputs (median MII at preferred delay = 0.33, IQR = 0.59, **Figure 3N**). Importantly, in a subset of experiments we studied if the order of presentation of auditory and visual stimuli affected nonlinear interactions. We found that supralinear summation was weaker when auditory stimuli preceded visual stimuli (median MII auditory leading = 0.02 ± 0.50, median MII visual leading = 0.15 ± 0.41; $P = 1 \times 10^{-4}$, Mann-Whitney U test; **Suppl. Figure 3**). This suggest that the temporal window for multisensory integration is asymmetric and might be best matched to represent visual information followed by auditory. Overall nonlinear multisensory interactions were observed throughout the different anatomical axes of the SC. The mean absolute MII, which accounts for supra- and sublinear interactions, was larger in the deeper layers of the SC (Linear mixed effects model, LMM, $P = 1 \times 10^{-7}$, ML: P = 0.9; AP: P = 0.2; DV: $P = 1 \times 10^{-8}$; **Suppl. Figure 4**). These results indicate that, as in other species[36,37], mouse SC neurons exhibit nonlinear summation of AV information[22,34].

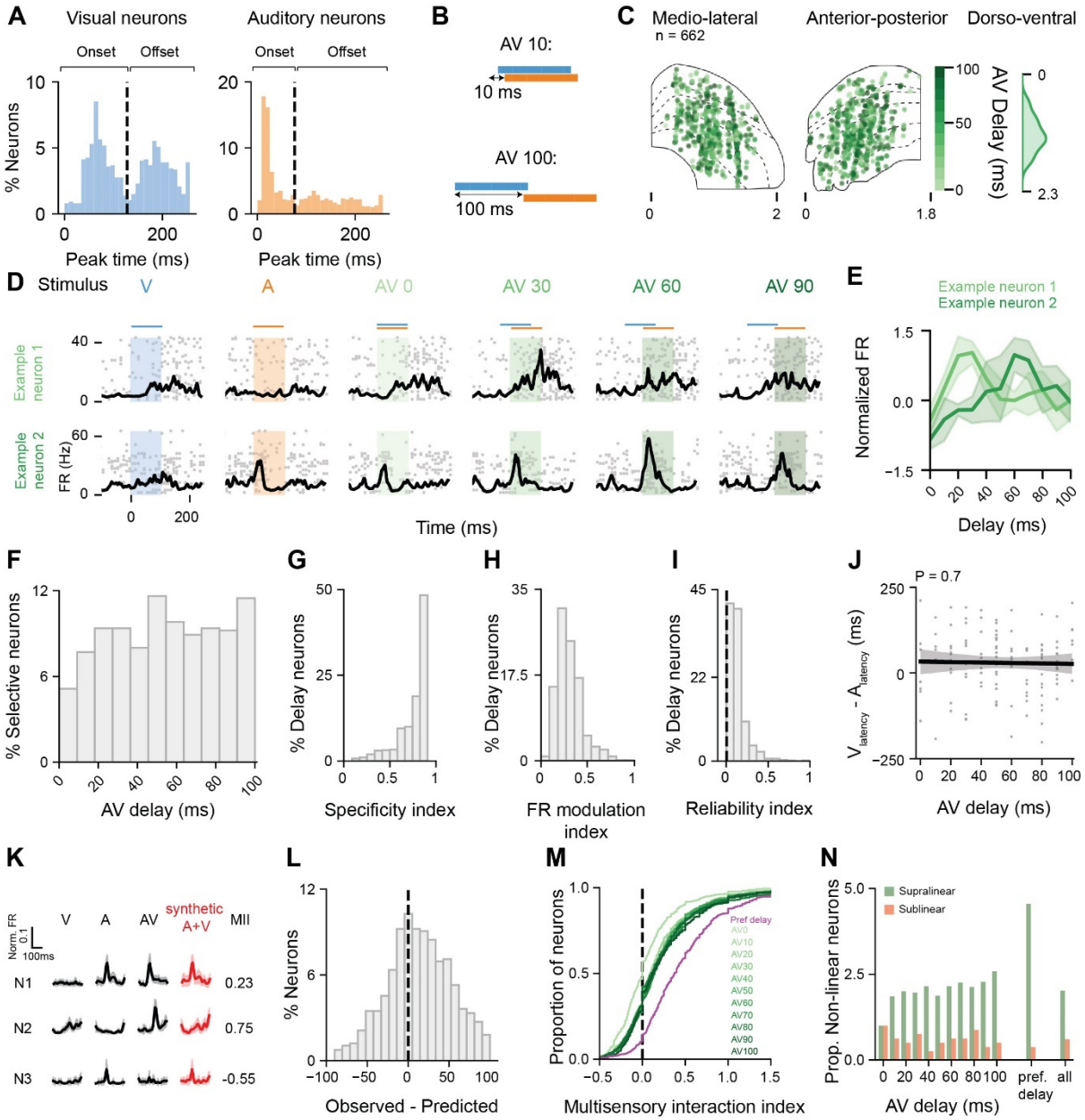


Figure 2 | Temporal delay of audiovisual information is encoded in superior colliculus neurons. A. Distribution of response latency for visual (blue left) and auditory (orange right) unisensory responses. **B**. Schematic of stimulation protocol. **C**. Anatomical distribution of neurons modulated by AV delay. Left: ML; Centre: AP; Right: DV axes. **D**. Raster plots and PSTHs for two example neurons to unisensory and multisensory stimuli presented at variable delays. **E**. AV delay tuning curve for neurons shown in **D**. **F**. Distribution of preferred AV delay across the population of delay-selective neurons. **G**. Correlation between the neurons' preferred AV delay and the difference in latency between visual and auditory response peak time. Only delay neurons that were responsive during both visual and auditory stimulation were included (N = 123). P value from linear regression. **H-J**. Distribution of specificity index (**H**); firing rate modulation (**I**) and reliability index (**J**) for delay-selective neurons. **K**. Unisensory and multisensory responses for three example neurons and corresponding synthetic multisensory response obtained from the addition of auditory and visual instantaneous firing rates. Numbers indicate multisensory interaction index (MII). **L**. Distribution of difference between observed and predicted AV delay preference for AV delay-selective neurons. **M**. Cumulative distribution of multisensory interaction index. Green tone indicates AV delay; magenta corresponds to MII calculated at the preferred delay of each individual neuron. **N**. Proportion of neurons exhibiting supra- and sublinear multisensory interaction, normalised by the number of neurons exhibiting supra- or sublinear multisensory interactions at AV delay 0 ms (supralinear = 8.9%; sublinear = 1.2%). Proportions are shown for each AV delay tested, as well for the preferred delay condition and as the mean value across all conditions (All).

Functional organisation of multisensory networks in the superior colliculus

To study if the functional network organisation of the SC could contribute to the nature of multisensory responses, we performed cross-correlation analysis between simultaneously recorded neuronal pairs along the DV axis[38–40]. This approach allowed us to study putative connections across the layers of the SC. To exclude stimulus evoked responses, we calculated the cross-correlogram (CCG)[41,42] on spike trains recorded during the inter-trial interval. Only responsive neurons were included in the analysis to avoid spurious correlations among unresponsive neurons (N = 3543 responsive neurons, 92 recordings, 24 mice). The approach used[39] identifies positive peaks in jittered-corrected CCGs, to correct for slow timescale correlations and excludes significant inhibitory correlations (see methods, **Suppl. Figure 5**). Putative connections were defined by lags of 1-5 ms, connection directionality was obtained from the sign of the lag, and the strength of the interaction was estimated from the peak of the CCG. Overall connection percentage was 2.3% (range: 0.2-10.7%, 2.3% vs false positive rate = 0.9%, P = 1 x 10^{-10}, Mann-Whitney U test) and decreased exponentially with increasing neuronal pair distance in the DV axis (**Figure 3A**). Interestingly DV connection percentage decayed along the ML axis of the SC with a higher percentage of connection for recordings from the medial SC compared to the lateral SC (ML: P = 1 x 10^{-5}; AP: P = 0.2, LMM; **Figure 3B**).

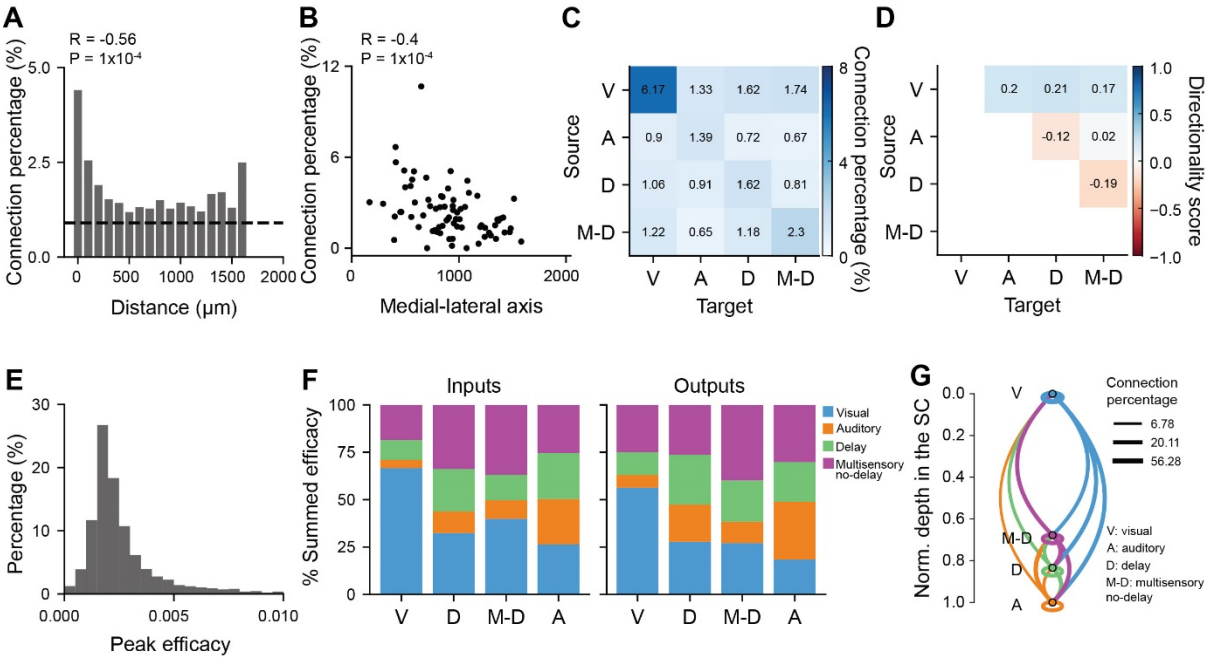


Figure 3 | Functional specificity of dorso-ventral connectivity within the SC. A. Mean putative connection percentage as a function of neuronal pair distance along the DV axis (step size 100 µm; N = 1421 putative connections over 69044 assessed, 92 recordings, 24 mice). Black dashed line represents the false positive rate estimated from applying cross correlation analysis to independently recorded neurons. **B**. Connection percentage as a function of mean location along the medio-lateral axis. Each dot represents one recording. **C**. Connectivity matrix between different subpopulations of neurons: V: visual, A: auditory, D: delay-selective, M-D: multisensory excluding delay selective neurons. **D**. Directionality matrix as in **C**. **E**. Distribution of excitatory putative connection strength. **F**. Percentage of the total summed efficacy for inputs (left) and outputs (right) for different functional subpopulations. **G**. Schematic representation of functional connectivity observed in the DV axes of the SC, based on **F**. Neuronal subpopulations are represented by circles, and are distributed according to their mean location along the DV axis of the SC. The line thickness indicated the total summed connection efficacy. Connections within functional subpopulations are shown as ellipses.

We then assessed connection probability and directionality between previously defined neuronal subpopulations (**Figure 3C,D**). This analysis revealed a high degree of functional specificity in the

network connectivity. The percentage of connections between visual neurons was 6.17%, while auditory neurons received the highest percentage of connections from other auditory neurons, albeit their connectivity rate was much lower (0.9%). Importantly, the remarkably higher connectivity of visual neurons compared to auditory neurons was not explained by a difference in firing rate during the inter-trial period used to estimate connectivity (**Suppl. Figure 5**). Overall, visual neurons provided most of the inputs to other neurons in the network, with the only exception of auditory neurons, consistent with the prominent visual role of the SC (**Figure 3C**). Interestingly, all multisensory neurons (AV delay-selective and delay-nonselective neurons) were more strongly connected with other multisensory neurons (**Figure 3C,D**) than to unisensory neurons. However, connectivity between delay-selective neurons was not explained by their preferred delay similarity (**Suppl. Figure 5**). Overall, the pattern of connectivity between functional subclasses was consistent across the ML axes of the SC albeit with a lower rate of connectivity in the lateral SC for all the functional subpopulations and in particular for delay-selective neurons (**Suppl. Figure 5**).

To further characterise network interactions, we used the interaction strength as a proxy for putative connection strength. The distribution of interaction strength was highly skewed (median interaction strength: 0.002; s.d.: 0.05; **Figure 3E**) with few strong interactions among a majority of weaker ones. To simultaneously account for interaction strength and interaction probability, we quantified the total sum of the interaction strength received by each neuronal functional subpopulation by combining data from all interacting pairs (**Figure 3F**). This analysis allowed us to build a model of putative DV connectivity of the SC (**Figure 3G**) and revealed that overall multisensory neurons (AV delay-selective and delay-nonselective neurons) receive less than 50% of the total local interaction strength from SC unisensory neurons, with the rest of the input arising from other multisensory neurons potentially contributing to their nonlinear multisensory integration.

Distinct contributions of SC subpopulations to the temporal representation of multisensory information.

The inter-trial variability (**Figure 2J**), and the relatively low response modulation evoked by specific AV delays in single SC neurons (**Figure 2I**), together with the widespread connectivity between multisensory neurons (**Figure 3F,G**) suggest that precise decoding of a wide range of AV delays rely on a distributed representation, i.e. from the collective activity pattern of the population. To assess how accurately AV delay can be predicted from population responses, we performed cross-validated, multi-class classification of single-trial responses using a linear classifier (support vector machine, SVM) and evaluated its performance on held-out data (80% training set, 20% test set). The classifier was trained to decode the 11 tested multisensory stimuli, as well as unisensory auditory and visual trials. We controlled for the lagged presentation of the auditory stimuli in multisensory trials by adding 10 synthetic unisensory auditory trials shifted in time (**Figure 4A**). This approach allowed us to distinguish the classifier's accuracy in decoding AV delays from its ability to determine the timing of a unisensory auditory response. The classifier achieved an 85% accuracy rate in decoding delay conditions when all neurons were included, outperforming chance levels (N = 5360 neurons; $P = 1 \times 10^{-8}$, Mann-Whitney U test, **Figure 4B,C**). Our analysis shows that decoding the range of AV delays used in this study can be achieved with as few as 9 neurons, resulting in a decoding accuracy significantly above chance level (> 9.09% decoding accuracy, $P = 1 \times 10^{-6}$, Wilcoxon signed-rank test, **Figure 4D**). Importantly we could not decode the AV delay from analysis of body movement (**Suppl. Figure 6**).

This result demonstrates that the SC can encode the precise temporal disparity between the onset of visual and auditory stimuli with a relatively low number of neurons. Furthermore, we observed a lower performance in classifiers trained on simultaneously recorded neurons compared to those trained on random samples of neuronal responses obtained across 92 recordings (N= 25 neurons per model; median accuracy for randomly sampled neurons = 19%, IQR = 10%; median accuracy for simultaneously recorded neurons = 11, IQR = 14; P = 0.003, Mann-Whitney U test, **Suppl. Figure 7**). This result suggests that noise correlations, amplified by recurrent connectivity, may constrain distributed population coding[43,44]. It might also indicate that variability on the representation of sensory information within the recorded population might contribute to the observed differences in performance[45].

To characterise the contribution of different neuronal populations to the decoding of AV delay, we compared the decoding accuracy of visual, auditory, multisensory delay-selective and multisensory delay-nonselective neurons, randomly sampling groups of 200 neurons from these populations. As before, auditory shifted trials were included as a control for each classifier. **Figure 4E** shows the confusion matrix for each group. This analysis revealed differences in AV delay decoding accuracy between subpopulations of neurons ($P = 1 \times 10^{-15}$, Kruskal-Wallis, **Figure 4F**). Auditory, multisensory delay-selective and multisensory delay-nonselective neurons showed the highest decoding accuracy (**Figure 4F**, auditory vs. visual neurons $P = 1 \times 10^{-8}$; delay-selective vs. visual neurons $P = 1 \times 10^{-8}$; delay-nonselective vs. visual neurons $P = 1 \times 10^{-8}$, Mann-Whitney U test, Benjiamini-Hochberg multiple comparison). Importantly, despite the high accuracy obtained for auditory neurons, this group also displayed a high rate of misclassification between AV delay trials and synthetic shifted auditory trials (**Figure 4G**). In contrast, the low decoding accuracy of visual neurons suggests that their responses are weakly modulated by the delayed presentation of an auditory stimulus. As expected, the highest accuracy was observed by multisensory neurons previously labelled as AV delay-selective (**Figure 4F**). The accuracy of the classification quantifies the rate of correct classifications. We next sought to quantify the precision of the classification for the different subpopulations. We therefore calculated the difference between the AV delay estimated from each trial classification and the actual AV delay for which the responses were measured. We quantified the classification precision of multisensory AV delay-selective and delay-nonselective neurons and found that for classifiers trained on the activity of 200 AV delay-selective neurons and tasked to decode AV delays ranging from 0 to 100 ms, separated by 10 ms intervals, the precision was ~ 2 ms (median ± IQR: AV delay-selective = 2 ± 1ms; multisensory AV delay-nonselective = 6 ± 2ms, $P = 1 \times 10^{-8}$, Kruskal-Wallis, **Figure 4H**).

To verify the contribution of different subpopulation to the decoding of AV delays within individual recordings, we analysed the decoding accuracy obtained from models trained on the responses of 25 randomly selected simultaneously recorded neurons using a linear mixed effects model (N = 70 recordings, 20 repetitions). High decoding accuracy was mostly attributed to the proportion of delay neurons present in the individual recordings and was not explained by the proportion of multisensory delay-nonselective neurons in the recording (multisensory delay-selective: $P = 1 \times 10^{-9}$, visual: $P = 0.03$, auditory: $P = 1 \times 10^{-6}$, multisensory delay-nonselective neurons: $P = 0.4$, linear mixed model (LMM), **Figure 4I,L**). Together these results highlight the contribution of AV delay-selective neurons to the accurate temporal representation of multisensory information.

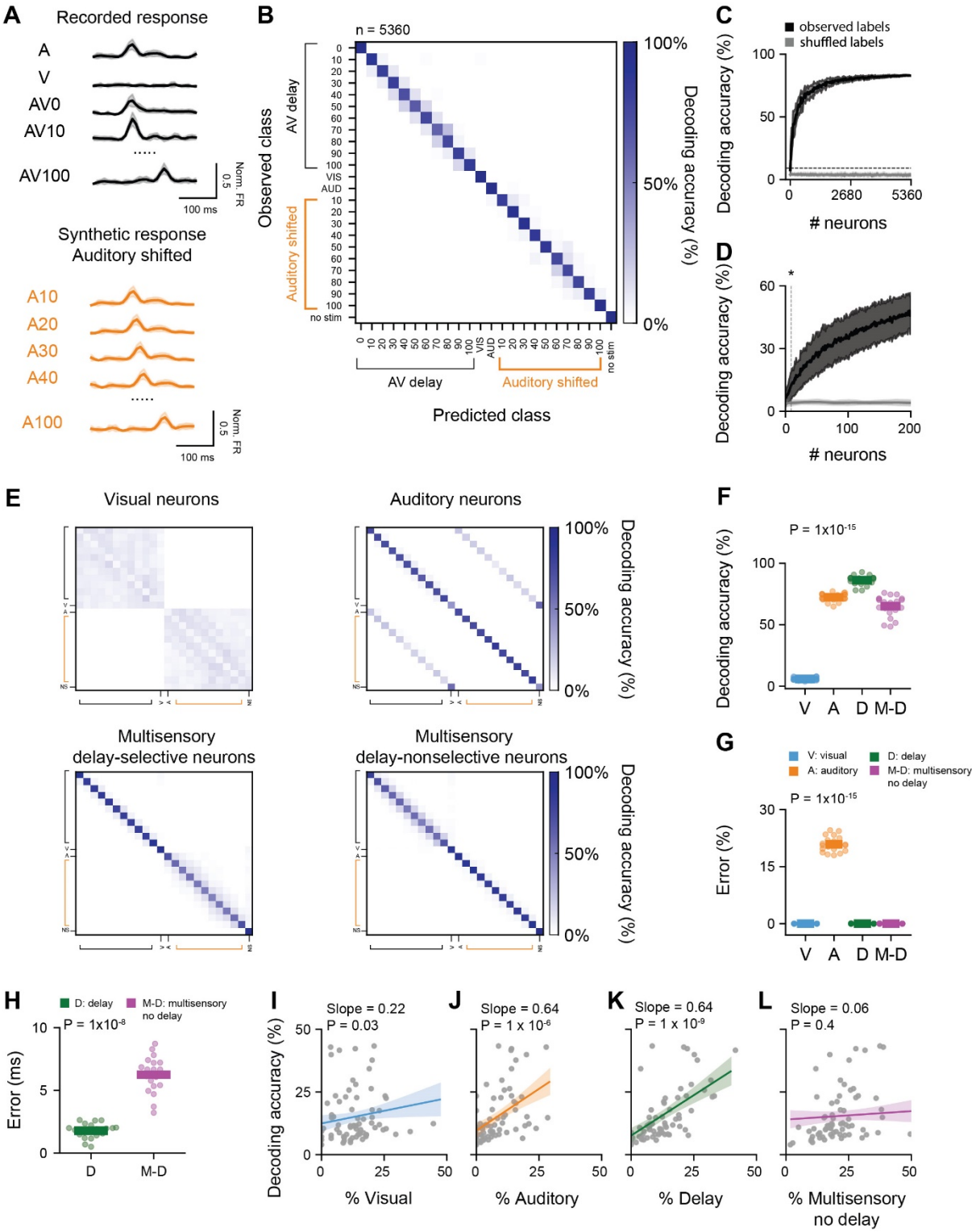

A
Recorded response
A
V
AV0
AV10
AV100
Norm. FR
0.5
100 ms
Synthetic response
Auditory shifted
A10
A20
A30
A40
A100
B
n = 5360
Observed class
AV delay
Auditory shifted
Predicted class
100%
50%
0%
Decoding accuracy (%)
C
observed labels
shuffled labels
Decoding accuracy (%)
neurons
D
Decoding accuracy (%)
neurons
E
Visual neurons
Auditory neurons
Multisensory
delay-selective neurons
Multisensory
delay-nonselective neurons
F
P = 1x10-15
Decoding accuracy (%)
V A D M-D
G
V: visual
A: auditory
D: delay
M-D: multisensory
no delay
P = 1x10-15
Error (%)
V A D M-D
H
D: delay
M-D: multisensory
no delay
P = 1x10-8
Error (ms)
D M-D
I
Slope = 0.22
P = 0.03
Decoding accuracy (%)
% Visual
J
Slope = 0.64
P = 1 x 10-6
% Auditory
K
Slope = 0.64
P = 1 x 10-9
% Delay
L
Slope = 0.06
P = 0.4
% Multisensory
no delay

Figure 4 | Population encoding of audiovisual delay in the superior colliculus. A. Example trials used to train a support vector machine (SVM) classifier. Neuronal responses aligned to the auditory stimulation during multisensory trials (orange traces) were included as a control for the classification of the timing of auditory stimulation. **B**. Confusion matrix obtained from linear classification of AV delay trials using 5360 neurons. Auditory trials shifted in time are included in the classification (orange labels). **C**. Mean decoding accuracy of linear classifier for AV delay trials as a function of number of neurons included in the classifier (black). The classifier trained on shuffled labels is shown in gray. The shaded area represents the standard deviation over 20 repetitions. The dotted line represents chance level. **D**. Decoding accuracy plotted as in **C**, for up to 200 neurons across 200 iterations. The dotted line represents the number of neurons necessary to achieve higher than chance performance. N = 9, P value from Wilcoxon signed-rank test). **E**. Confusion matrix for linear classification of AV delay trials for 200 neurons randomly sampled from different subpopulations: visual (V), auditory (A), delay-selective (D) and multisensory non delay-selective (M-D) neurons. Auditory trials shifted in time (orange labels) were included in the classification as in **B**. **F**. Mean decoding accuracy calculated from the confusion matrices in **E**. P value from Kruskal-Wallis test. **G**. Misclassification error rate (%) calculated for 200 randomly sampled neurons from different neuronal subpopulations. **H**. Misclassification error (ms) for 200 randomly sampled D and M-D neurons for classifiers only trained in AV delay trials. **I-L**. Linear decoders were obtained from individual recordings with at least 25 neurons. Partial dependence plots show mean classifier decoding accuracy as a function of the percentage of visual (**I**), auditory (**J**), delay-selective (**K**) and multisensory delay-nonselective (**L**) neurons. The bold line represents the fitted regression, while the shaded gray area denotes the 95% confidence interval. Dots indicate values from individual recordings, P values were calculated using a linear mixed model with mouse ID as random effect.

Nonlinear summation of auditory and visual inputs improves the temporal precision of multisensory representations.

AV delay selective neurons exhibit strong nonlinear integration of audio-visual signals (**Figure 2N**). We next sough to determine the extent to which nonlinear summation of sensory information contributes to the discrimination of AV delays at the population level. We compared two independent linear classifiers: one trained on the measured multisensory responses and the other on synthetic AV responses obtained by linearly adding unisensory visual and auditory responses, with the latter shifted by each of the temporal delays tested experimentally. As before, unisensory visual, auditory and auditory-shifted trials, along with a blank trial, were included as a control in both classifiers. When considering multisensory delay-selective neurons, the classifier trained on the measured multisensory responses outperformed the classifier trained on the synthetic linear responses (performance for 200-neuron classifier on observed nonlinear responses = 86 ± 4% vs synthetic linear responses = 6 ± 2%, P = 1 x 10^{-8}, Mann-Whitney U test, **Figure 5A**). Similar results were obtained when including a sample of neurons from the whole population, suggesting that small nonlinearities found across the population can also contribute to the decoding of multisensory information (performance for 200-neuron classifier on observed nonlinear responses = 46 ± 10% vs synthetic linear responses = 7 ± 4%, P = 1 x 10−8, Mann-Whitney U test, **Figure 5A**). Nonlinearities similarly increased accuracy when training a Random Forest Classifier, even when this classifier should be able to capture other nonlinear interactions in the population response (RFC; performance for 200-neuron classifier on observed nonlinear responses: 39 ± 13% vs synthetic linear responses 15 ± 9%, P = 1 x 10^{-6}, Mann-Whitney U test, **Figure 5A**). Similar results were obtained when comparing the performance of classifiers trained on individual recordings (**Figure 5B**). In this case, the whole population was included given the small sample size found on individual recordings. Moreover, decoding accuracy on individual recordings was strongly correlated with the degree of nonlinearities in the recorded population (P = 1 x 10^{-3}, **Figure 5C**). These results demonstrate that nonlinearities in SC neuronal responses enhance the accuracy of the representation of AV information.

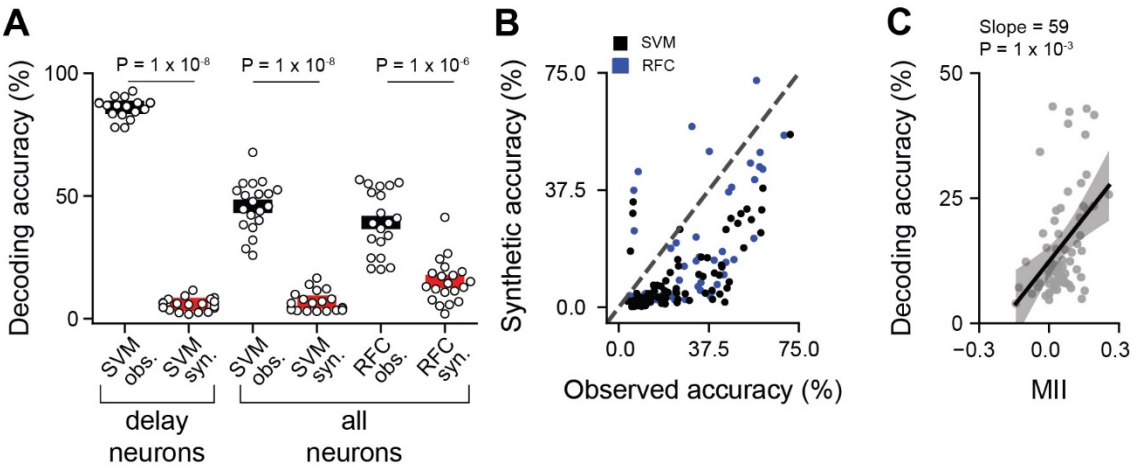


Figure 5 | Encoding of audiovisual delays is enhanced by nonlinear summation of unisensory inputs. A. Comparison of decoding accuracy between classifier trained on recorded responses (black) and classifier trained on synthetic responses (red). The classifiers were trained on the responses from 200 delay selective neurons (left) or randomly sample neuronal populations (right). P values from Wilcoxon signed-rank test. **B**. Decoding accuracy for classifier trained on synthetic responses versus accuracy from classifier trained on observed responses, for individual recordings. Number of neurons included in the classifier varied for each recording. Range 11-130 neurons per recordings; N = 92 recordings. Black and blue circles correspond to SVM and RFC, respectively. **C**. AV delay decoding accuracy as a function of the mean MII of all neurons within individual recordings. The bold line represents the fitted regression line, while the shaded gray area denotes the 95% confidence interval. Gray circles correspond to individual recordings. P value from linear mixed model.

Enhanced audiovisual delay encoding in the medial - posterior SC.

Previous studies have shown asymmetries in behavioural multisensory interactions attributed to the SC, with stronger audiovisual integration in the temporal hemifield compared to the nasal hemifield[46] (Bertini et al., 2008). Functionally distinct regions of the SC exhibit unique patterns of inputs and outputs[27], as well as differential connectivity along the DV axis (**Figure 3B**) that could give rise to differences in multisensory processing across the sensory field. These distinct local properties could explain the variability in AV delay encoding accuracy observed across individual recordings (**Figure 5C)**. Therefore, to investigate how population-level encoding of AV delays varies along the ML, AP and DV axes of the SC, we trained the SVM classifier on individual recordings and assessed how AV delay decoding accuracy changed with the different axes within the SC. We found that decoding accuracy exhibits a gradient along the ML and the AP axes (N = 70 recordings, 25 neurons per recording, P = 1 x 10^{-3} multivariate linear regression, **Figure 6A,B**). Similar results were obtained when training the SVM on randomly sampled neurons within spatial bins of 650 µm along each axis. Decoding accuracy varied along the three anatomical axes (P value: ML = 0.02, AP = 1 x 10^{-5}, DV = 1 x 10^{-4}; Kruskal-Wallis test, **Suppl. Figure 8**). Specifically, accuracy in the most medial portion of the SC was, on average, 11% higher than in the most lateral region (P = 0.03, Mann-Whitney U test, Benjamini-Hochberg multiple comparison, **Suppl. Figure 8**). Similarly, accuracy in the posterior SC was 12% higher than in the anterior SC, and along the DV axis the deepest SC region showed 12% higher accuracy (P value AP = 1 x 10^{-4}, DV = 1 x 10^{-3}, Mann-Whitney U test, Benjiamini-Hochberg multiple comparison, **Suppl. Figure 8**). These findings were robust across a range of bin sizes and neuron sample sizes (**Suppl. Figure 8**). When considering all three anatomical axes simultaneously by randomly sampling 50 neurons from eight distinct areas of the SC, each covering a volume of 1.09 mm^3 (528 ± 243 neurons per bin), the accuracy remained higher in the posteromedial region (P = 1 x 10^{-16}, Kruskal-Wallis, **Suppl. Figure 8**).

To understand if the difference in decoding accuracy could be explained by a differential representation of neuronal subpopulations, we compared the proportion of visual, auditory, multisensory delay-selective and delay-nonselective neurons across the different anatomical axes of the SC. The SC was divided into four subregions, each covering 2.20 mm^3 (1039 ± 121 neurons per bin), and the distribution of neuronal subpopulations was compared across these regions. Our analysis revealed no significant correlation between decoding accuracy and the distribution of subpopulations (for percentage of visual neurons: P = 0.4, auditory: P = 0.06, delay: P = 0.1, multisensory delay-nonselective neurons: P = 0.1, LMM, **Figure 6C**). Additionally, the preferred delays of AV delay-selective neurons showed no

differences across spatial bins (P > 0.4 paired comparisons across all subregion combinations, Kolmogorov-Smirnov test, **Figure 6D**). Moreover, the observed regional encoding differences were maintained when training classifiers on subpopulations of visual, auditory, multisensory and delay-selective neurons, although decoding accuracy was similar for the latter group across the ML axis (**Suppl. Figure 9**).

We next examined if the degree of nonlinear integration of multisensory information across anatomical regions might explain the gradient in decoding performance. The difference in AV delay population decoding across SC regions was not explained by differences in the averaged multisensory integration index across the population in specific SC subregions (P = 0.7, LMM). However, the mean multisensory index does not capture the diversity of multisensory interactions across the population. We therefore studied if nonlinear integration might contribute to the observed gradient in decoding of AV delays by training classifiers on synthetic nonlinear responses. The decoding accuracy decreased across all SC regions when training the SVM classifier with synthetic linear responses (**Figure 6E**) and although a small gradient in decoding performance was still present, the decoding accuracy was not higher than chance level for any of the bins tested.

Taken together these results show that SC regions encoding information from the upper-temporal visual field are best suited for representing temporal asynchronies in the onset of visual and auditory signals. This functional specialisation of multisensory temporal integration in SC posteromedial subregions likely arises from nonlinear integration of AV signals in mixed functional subpopulations specifically interconnected.

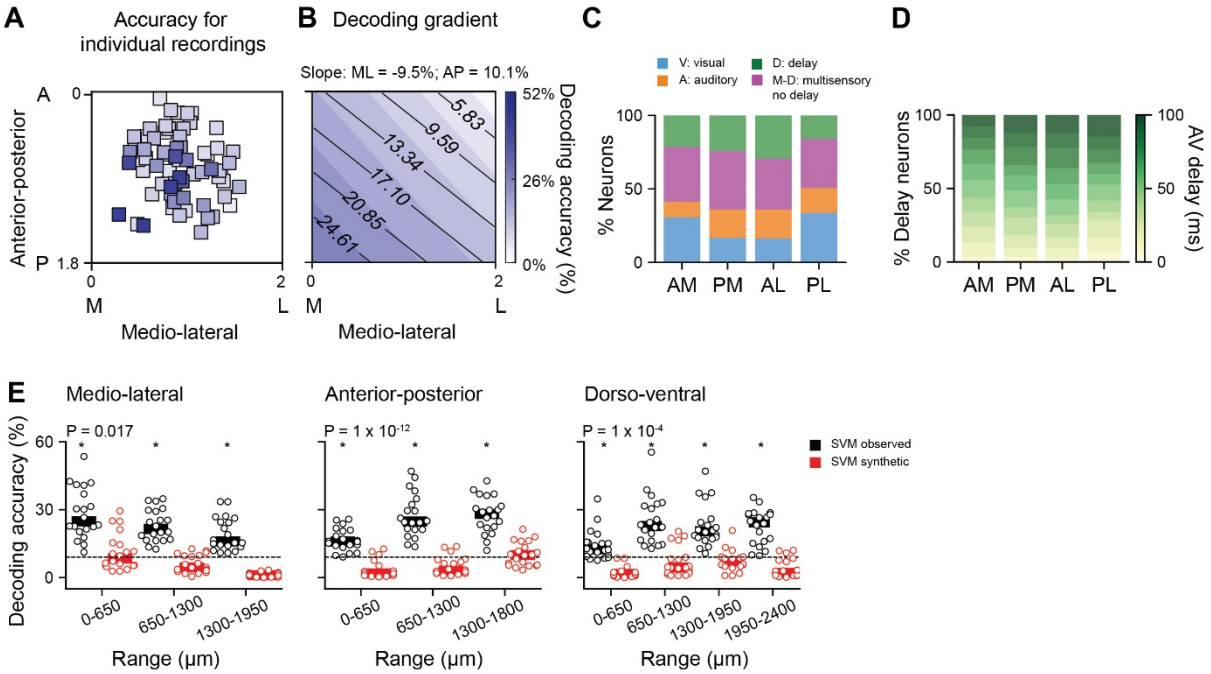


Figure 3 | Population decoding of audiovisual delays is enhanced in the posteromedial SC. A. Anatomical distribution of individual recordings colour coded according to the mean AV delay classifier accuracy obtained with 25 neurons, 20 repetitions. **B**. Decoding gradient as a function of ML and AP location, calculated from the results shown in **A. C**. Percentage of visual (V), auditory (A), delay-selective (D) and multisensory non delay-selective (M-D) neurons across four anatomical bins: AM: anterior-medial; PM: posterior-medial; AL: anterior-lateral; PL: posterior-lateral. Each bin covered 2.20 mm^3, with on average 1039 ± 121 neurons per bin. **D**. Percentage of neurons selective for different AV delays (0-100 ms) for the same anatomical bins as in panel (**C**). **E.** AV delay decoding accuracy for 20 iterations of the linear classifier trained on the responses of 50 randomly sampled neurons within 650 µm bins along the ML (left), AP (centre) and DV (right) axes of the SC. The classifiers were either trained on the observed responses (black) or on temporally shifted linearly summed audio visual responses (red). The dotted line represents chance level. *= P < 0.05 with respect to chance, Wilcoxon signed-rank test. P values indicate differences across anatomical locations for observed responses, Mann-Whitney U test.

Discussion

We investigated the temporal integration of auditory and visual information in the mouse SC at the single-neuron and population levels. A wide scope of previous work has shown that multisensory integration across brain regions depends on the temporal relationship of the sensory inputs[47–49]. We found neurons exhibiting tuning to AV asynchronies that are widely distributed across the SC. The AV delay preference of individual neurons emerges from nonlinear integration of unisensory inputs, resulting in delay preferences longer than those predicted by linear summation of unisensory responses. These nonlinearities are more prevalent when visual information precedes auditory input, suggesting a mechanism that supports the efficient encoding of ethologically relevant multisensory information. Using population-level decoding via SVM classifiers, we show that nonlinear integration enhances the precision and accuracy of the SC's representation of AV delays. Furthermore, the medial-posterior SC, which represents the peripheral visual field, exhibited greater AV delay encoding accuracy compared to lateral and anterior regions of the SC. Functional connectivity analysis using cross-correlation revealed that SC neurons tend to form connections within their functional subclasses—visual, auditory, multisensory, or AV delay-selective neurons. This functional specificity suggests that SC connectivity is organised to optimise sensory integration while preserving the distinct processing roles of these subclasses.

Nonlinear audiovisual integration in the mouse SC

Nonlinear AV integration in the mouse SC has been previously reported, with early auditory processing modulating later visual responses as a proposed mechanism for multisensory integration[22]. However, recent reports failed to find spatial auditory RFs and nonlinear summation of AV information in the mouse SC[50]. The authors suggested strong modulation of auditory neurons by body movements[50], highlighting the challenges of measuring auditory responses in behaving animals. To minimise the effects of uninstructed stimulus-evoked movements in our study, we recorded from habituated and passively stimulated mice using low sound intensity. The auditory origin of our recorded responses is supported by the short latency of auditory neurons, the specificity of AV delay neurons and the absence of AV delay-specific triggering of uninstructed movements. Our approach enabled us to identify auditory and multisensory RFs that were spatially aligned to visual RFs of neurons recorded within the same anatomical location. We characterised AV integration in the temporal domain and identified AV delay-selective neurons, a subset of multisensory neurons that exhibit nonlinear summation of sensory inputs at the tested temporal disparities. Such nonlinear selectivity can increase the dimensionality of sensory representations, enhance stimulus discriminability, minimise errors, and provide a robust neural code[51,52]. Consistently, our results show that nonlinear multisensory integration improves the accuracy of AV delay decoding compared to linear models. This mechanism may facilitate the extraction of population responses by downstream regions[51], such as the lateral geniculate nucleus (LGN), lateral posterior nucleus of the thalamus (LP), zona incerta (ZI) and the parabigeminal nucleus (PBG)[28,53], to support SC-mediated sensory and motor responses[54].

The cellular and circuit mechanism underlying multisensory responses in SC neurons are not yet fully understood. Modelling studies suggest that clustering of synaptic inputs may play a key role[55–57]. Indeed, multisensory responses in collicular neurons depend on NMDA-gated channels[58], suggesting that the spatial arrangement of synaptic inputs could contribute to nonlinear local integration. This may lead to computational capabilities such as directionally biased responses, proximal- distal synapse interplay, and classical RF–contextual interactions[59–61]. Intrinsic membrane properties[62] and inhibitory mechanisms[23,49,55,62–64] have also been proposed to explain multisensory responses of SC neurons. Future studies at single-neuron level[65–68] should explore how the functional arrangement of synaptic inputs contributes to multisensory integration, while research on long range connectivity should address the role of cortical input[57,69]. A genetic market for identifying multisensory neurons would be a valuable tool for this purpose.

Ethological relevance of audiovisual delay representation in the SC.

The observed tuning of SC neurons to AV delays has clear ethological relevance. Auditory signals typically reach the SC ~50 ms earlier than visual signals [22,70] due to differences in receptor latencies—ranging from 20–100 ms in the retina[71] to 1–4 ms in the cochlea[72]. An optimal integration window for

stimuli with a 50 ms asynchrony is consistent with the timing that evokes optimal perceptual coincidence[73,74]. However, SC neurons are tuned to a broad range of AV delays[23,24]. This broad tuning might support the binding of cross-modal stimuli despite differences in their arrival times, by providing a buffer for latency differences between sensory modalities[75]. The consistent difference in the speed of light and sound introduces a regular feature in multisensory signals: visual input consistently precedes auditory input in a distance-dependent manner, counterbalancing the different latency of the two sensory systems. Experience-dependent development of multisensory temporal windows aligns sensory integration with the statistics of natural events [74], and broad AV delay selectivity may enable robust detection across varying target distances. Our observation that nonlinear responses are overrepresented when visual signals precede auditory signals suggests that this mechanism might allow the efficient representation of the temporal features from the natural environment. It would also enable robust detection of external events even at low contrast visual input or low sound intensity which evoke neuronal responses with high latency variability. This adaptive strategy ensures animals can respond effectively to dynamic and variable environmental stimuli. Analogous nonlinear mechanisms have been observed in the encoding of target distance through echo delay representation in bats [76,77], suggesting a conserved function across species for processing temporal disparities.

Non-uniform representation of AV delays in the SC.

Our study revealed higher discriminability of AV delays in the medial-posterior SC, which encodes the upper temporal sensory field. Biased representation toward the temporal hemifield has been observed in SC-mediated behaviours, such as saccadic movement[78], covert orienting behaviour[79], unconscious priming[80] and saccadic latencies[81]. Biases along the anterior-posterior axis of the SC might be attributed to the non-homogenous representation of the visual field due to asymmetries in retinal projections[53,82–85]. Multisensory interactions also exhibit an asymmetric expression in humans, where AV integration is stronger in the temporal hemifield than the nasal hemifield[46]. Eccentricity biases of multisensory interactions have also been observed in human cortical visual areas, where decoding of auditory information improves toward the peripheral retinotopic regions in early visual cortex (V1, V2, and V3)[86,87]. This eccentricity-based functional organisation is predicted by primate anatomical connectivity[88], where feedback projections from non-visual areas preferentially target the periphery of early visual areas. This differential multisensory representation may enhance reactions to peripheral, unattended stimuli — a critical function for survival.

In rodents, the medial SC is crucial for innate defensive responses to overhead looming stimuli[15] and plays a key role in escape behaviours[16]. Both the medial and posterior SC are involved in turning behaviour[12,89], achieved through a highly interconnected network with cortical and subcortical structures associated with visual and auditory processing, and visuomotor responses[28]. These connections support the SC's crucial role in guiding spatial navigation[90,91], visual attention[90,92] and orienting behaviour[12,93], which underscores the importance of achieving a precise representation of AV temporal discrepancies for accurately guiding animals in their behavioural responses. An accurate estimation of AV delays could serve as a mechanism for predator's distance estimation and therefore guide the rapid selection of defensive responses, such as freezing or escaping.

Functional specificity of SC connectivity

Putative connectivity was assessed through cross-correlation analysis of inter-trial-interval responses across the SC's dorsoventral axis. Our analysis indicates that connections within functional subclasses are overrepresented, with the most prevalent connections occurring between visual neurons. Interestingly, multisensory neurons receive approximately 50% of functional local inputs from multisensory neurons and the remaining from visual and auditory neurons, suggesting a complex interconnected network. Recurrent connectivity between multisensory neurons could extend the temporal window of integration and explain the longer AV delay selectivity than the predicted by the linear summation of unisensory responses. The overall rate of connectivity decreased from the medial SC to the lateral SC, with delay-selective neurons displaying the lowest rate of connectivity in the lateral SC (**Suppl. Figure 4**) consistent with lower AV delay encoding in lateral regions. Our jitter correction and high threshold over the noise distribution for estimating connectivity likely resulted in an

underestimation of connection probabilities. Future work using multiple probes[94] should assess lateral connectivity potentially contributing to cross-modal interactions in the spatial domain. Furthermore, circuit mapping through viral tracing[95,96] or electrophysiological methods in slices[97] would be needed to confirm our results. However, previous work has shown that correlative analysis of neuronal responses measured *in vivo* can provide good estimates about synaptic connectivity measured in brain slices of the same tissue[98]. A highly interconnected network of multisensory neurons might provide a mechanism for extending the multisensory binding window, ensuring robust and precise encoding of sensory signals. This ensures reliable detection of targets across a wide range of environmental conditions, while nonlinear integration enhances the precision of temporal information encoding.

Methods

Animals

All animal procedures performed in this study were licensed by the UK Home Office and approved by the Crick Institutional Animal Welfare Ethical Review Panel. Experiments in this study were performed in 31 male and female C57BL/6J.Cdh23$^{753A>G}$ mice (MRC Harwell, UK)[99], aged 2–4 months. This mouse line has a CRISPR/Cas9- mediated homology repair of allele Cdh23 that prevents age-related hearing loss[99]. Mice were housed in a 12 h/12 h dark/light cycle with ad libitum access to food and water. All experiments were carried out in sound-attenuated chambers. Mice were randomly assigned to the different experimental groups.

Surgical procedures

Mice underwent two surgeries (metal headplate implantation and craniotomy). Analgesia was provided on the day prior to surgery (Metacam 5mg/kg in custard). For both surgeries, mice were anesthetised with isoflurane (5% induction, 2-2.5% maintenance) and injected subcutaneously with analgesic and anti-inflammatory compound (10 mg/kg Meloxicam and 0.1 mg/kg Buprenorphine, dosage of injectable 0.1 ml/10g of body weight). The animals body temperature was maintained at 37C° with a thermal mat. A metal headplate was fixed to the skull using dental cement (Super Bond C&B, Sun Medical) centred on the right SC (AP: -3.7 mm and ML: 0.9 mm from bregma). After up to 2 weeks of habituation to the behavioural apparatus, animals underwent a second surgery to perform a craniotomy of 2-3 mm in diameter and a durectomy. The craniotomy was covered with artificial dura (Dura-Gel, Cambridge NeuroTech) and sealed with Kwik-Cast silicone elastomer (World Precision Instruments). Mice were left to recover for at least 24 hours before recordings.

Neuropixels recordings

After recovery (minimum 24h), neural activity was recorded using Neuropixels 1.0[29]. For 2-7 days Neuropixels recordings were performed on awake, head-restrained animals. Each animal underwent a maximum of 5 probe insertions, with no more than two insertions per day. The order of the anatomical location of the insertion was randomised. By varying the position of the recording site, neuronal activity was acquired along the medial-lateral axis (ML: 0 mm to 1.9 mm from midline) and anterior-posterior axis (AP: -3.08 mm to -4.72 mm from bregma) portion of the colliculus. Neuropixels recordings were performed using a National Instruments I/O PXIe-6341 module and OpenEphys software (https://github.com/open-ephys). The probe was coated with DiI (ThermoFisher Scientific, V22885) or DiO (ThermoFisher Scientific, V22886) on the day of the recording, to be able to validate recording location. On the last day of recordings, the animals were perfused and the extracted brains were imaged with serial two-photon (STP) tomography[100] to perform post-hoc tracing of the electrode locations. Brainreg was used to register the brains to a reference atlas[101] and the probes were traced with brainreg-segment[101].

During the recording, animals were 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). Additionally, in a subset of experiments, to record uninstructed body movements during sensory stimulation, videos of the face of the animals were acquired through an CMOS camera (Basler

acA1440-220um), coupled with a 50 mm lens (Computar Lens M5018MP2). Videos were recorded at 200 Hz and the acquisition was synchronised with Neuropixels recording using a common TTL trigger sent by the HARP RGB LED board.

Sessions were automatically spike-sorted using Kilosort2 (https://github.com/MouseLand/Kilosort/releases/tag/v2.0) and manually curated to select isolated single cells using Phy (https://github.com/cortex-lab/phy). Units were classified as 'good' based on the following criteria: they exhibited a strong refractory period in the cross-correlogram, had a typical waveform, and maintained a stable firing rate throughout most of the recording. Recordings were excluded from further analysis if, upon visual inspection of the spike-sorted data and anatomical reconstruction of the probe track, the anatomical boundaries of the SC could not be determined.

Sample size was estimated based on the expected effect size based on similar studies[14,22,33] and the current standard in mouse neuroscience studies.

Stimuli

Mice were presented with visual, auditory and multisensory stimuli generated using Psychophysics Toolbox (www.psychotoolbox.com) in Matlab. To present visual and auditory stimuli simultaneously at corresponding locations on the opposite side of the recorded SC, we designed a specialised device using 0.8mm thick laser cut PCBs in collaboration with the *Making lab*, a technology platform from the Francis Crick Institute. The device we created consisted of a hemispheric array with seventy-five stimulation positions, each housing a speaker and an LED. These stimulation positions were located 18° apart in both azimuth (horizontal) and elevation (vertical) directions. As a result, the arrangement formed a spherical section spanning 180° in azimuth and 72° in elevation, with a radius measuring 18 cm. Details of the stimulation apparatus can be found at (https://github.com/FrancisCrickInstitute/Coliseum).

Visual stimuli were presented via a dimmable LED (2.4 lux, 4.3°, OSRAM SmartLED, LWL283-Q1R2-3K8L-1) and auditory stimuli consisting of white noise (1-22 kHz, 60-65 dB intensity) were presented via a speaker (VISATON, K 28 WPC BL). Light diffusers (1.5 cm diameter thick Formlabs clear resin V4, FLGPCL04) were positioned in front of each LEDs. For receptive field mapping visual and auditory stimuli were presented independently or simultaneously in 5 elevations (-36° to 36° with 18° steps) and 7 azimuths (between -162° to 0° with 18° steps), totalling 35 locations with 25 repetitions per stimulus.

To study the effect of AV temporal synchrony, we first determined the centre of the receptive field of the neurons in the contralateral side of the recording site, and then presented visual and auditory stimuli either alone or combined. When together, two combinations were tested: visual-leading and auditory-leading stimuli. In the visual-leading condition, the auditory stimulus was presented with a delay between 0 and 100 ms relative to the visual stimulus, with 10 ms increments. In the auditory-leading condition, the visual stimulus was delayed by 0 to 100 ms relative to the auditory stimulus, with 25 ms increments. In both protocols, stimulus duration was 100 ms and the number of repetitions was 50. Interspersed trials with no stimulation were used as within subject control. In all experiments the stimulus sequence was randomised in a block design, all the stimulation conditions were randomised, and this was repeated the number of times previously specified per protocol.

Data analysis

All data analysis was done in MATLAB 2022b or Python. Data analysis was performed blind to the anatomical location of the recording sites.

Classification of neurons for AV delay protocol

A total of 5360 neurons (92 recordings, 24 animals) were used for the analysis. Neurons were classified based on their response to different stimuli using the ZETA test[35]. To account for multiple comparisons, p-values were corrected using the Benjamini-Hochberg procedure for the number of stimuli. Neurons were categorised as visual-responsive if they showed a significant response solely to visual stimulation.

Similarly, auditory neurons were defined as those that exhibited a significant response exclusively to auditory stimulation. Neurons were considered multisensory if their response pattern fell into one of the following categories: Bimodal – neurons that responded to both visual and auditory stimulation; Gated – neurons that exclusively responded to multisensory stimulation.

Receptive field analysis

A total of 979 neurons (11 recordings, 9 animals) were used for this analysis. To estimate the RF of individual neurons, a 2-dimensional Gaussian was fitted to the peak firing rate (FR) response for each stimulus. We computed Pearson's correlation between the location of the centre of the mean visual RF and the mean auditory RF for individual AV neurons (n = 89 neurons) and for each recording (N = 11 recordings). Moreover, we analysed the spatial alignment of RFs at the population level (N = 11 recordings).

Unisensory latency estimation

The spiking rate of individual unisensory neurons was binned in 1 ms bins during unisensory stimulation. The latency to peak for each neuron and condition was determined as the time needed to reach the maximum FR of the response profile. Onset and offset time windows were estimated based on the latency data. Gaussian kernel density estimation was applied separately to the distributions of latencies corresponding to the stimuli. This approach facilitated the identification of the underlying bimodal distribution of peak latencies, indicative of distinct response phases. The onset and offset windows for visual latencies where from 0-128 ms and 129-250 ms respectively. Onset and offset windows for auditory latencies where from 0-76 ms and 77-250 ms respectively.

Delay tuning and selectivity indices

To assess if the AV delay multisensory conditions differed from the highest unisensory condition, we tested the null hypothesis that peak firing rate to the multisensory and unisensory trials do not differ. To simulate data that would be observed under the null hypothesis, multisensory and unisensory trials were merged and then split randomly into two groups. We calculated the peak FR of such two groups by computing the mean FR rate over a 20 ms window centred around the maximum response time. The difference peak FR for the two groups was then computed. A bootstrap distribution of peak differences was then generated by 10000 repetitions. The difference value obtained from the actual multisensory and unisensory groups was compared to the 2.5th and 97.5th percentiles of the bootstrap distribution. If the actual difference fell outside this range, it suggested that the null hypothesis could be rejected and that there was a significant difference between the conditions. After identifying the AV delay conditions that differed significantly from the highest unisensory condition, the preferred delay was determined as the one that elicited the maximum FR among the AV delays showing significant differences from the preferred unisensory modality.

The delay selectivity of the neurons was quantified using three measures: the specificity index, the FR modulation index and the reliability index.

The specificity index (SI) evaluates the proportion of stimuli that significantly exceed the unisensory stimulation over the total number of stimuli presented, as is quantified as follows:

$$SI = 1 - \frac{sig_{delays}}{N_{delays}} \qquad \text{Eq. 1}$$

A SI value close to 1 indicates high specificity, with neurons responding uniquely to one AV delay condition. Conversely, a SI close to 0 suggests that the neuron exhibits low selectivity to individual delays, by responding to more AV delay stimuli.

The FR modulation index measures the change in FR between the preferred AV delay condition and the least effective AV delay. It was quantified as follows:

$$FR_{modulation\ index} = \frac{AV_{max} - AV_{min}}{AV_{max} + AV_{min}} \qquad \text{Eq. 2}$$

Where AV_{max} and AV_{min} are the mean peak FR over a 20 ms window for the preferred AV delay and the AV delay evoking the lowest firing rate, respectively. Higher values of this index indicate a greater change in FR between the preferred and least effective AV delays.

The reliability index (RI) corresponds to the mean Pearson's correlation coefficient for each trial's repetitions within the same trial type.

$$RI = \frac{2}{R(R-1)} \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} r(x_i, x_j) \qquad \text{Eq. 3}$$

where R is the number of repetitions, $r(x_i,x_j)$ is the Pearson's correlation coefficient between the responses at the ith and j^{th} repetitions and x_i,x_j represent the response at repetition i and j respectively. Higher RI values indicate high reliability of neurons and low trial-to-trial variability. To determine the chance level reliability, the RI was calculated for randomly selected time intervals within the total spike train, where responses were sampled independently of the trial type.

Multisensory interaction index

To measure the deviation of multisensory responses to the sum of unisensory responses a multisensory interaction index (MII) was computed as follows:

$$\text{MII} = \frac{\text{AV} - (\text{A+V})}{(\text{A+V})} \qquad \text{Eq. 4}$$

Were AV constituting the peak FR of the responses to audiovisual stimulation whereas A+V responses represent the peak FR of post hoc summed auditory and visually evoked unisensory responses. To generate A+V delay trials, the recorded auditory responses were shifted according to the delay of the auditory stimuli during the corresponding AV delay trials. The mean baseline response was subtracted to account for the shift in baseline in the summed responses. Nonlinear interactions were identified when there was a 50% change in FR between the observed AV response and the sum of unisensory responses. Positive MII values indicate supralinear multisensory interactions, whereas negative MII values indicate sublinear interactions.

Spatial correlation analysis

To analyse the relationship between the neurons' spatial location and their MII, a linear mixed model was used. The model included the animal ID of each neuron as a random effect with a varying intercept and a fixed slope. The MATLAB function '*fitlme*' was used, resulting in the following notation:

$$MII \sim 1 + ML_{position} + AP_{position} + DV_{position} + (1|animal\ ID) \qquad \text{Eq. 5}$$

were $ML_{position}$, $AP_{position}$, and $DV_{position}$ correspond to the position of each neuron along the medio-lateral, anterior-posterior and dorso-ventral axes, respectively.

Population analysis

A linear classifier (support vector machine, SVM) and a nonlinear classifier (random forest classifier, RFC) were used to estimate population delay decoding accuracy. All neurons were included in the decoder analysis. The spiking rate of individual neurons, binned in 10 ms bins during single trial stimulation, was used to decode the AV delay trials. Classifiers were trained with increasing number of neurons, reaching the maximum neuron's number in the population (5360) or the maximum neuron's number on individual recordings, when trained on single recordings.

Two main classifiers were trained. A first classifier was trained to estimate decoding accuracy between audio-visual trials and auditory shifted trials were included to account for the classification of delayed

auditory responses. Auditory shifted trials were obtained by aligning the auditory response to the onset of auditory stimulation during AV delay trials.

All classifiers were trained on 80% of data and assessed on 20% held-out data. We used a 5-fold cross validation and averaged the decoding accuracy of 5 runs. The decoding accuracy reported corresponds to the classification accuracy for AV delay trials only. The significance of the decoding accuracy (compared to chance) was computed by performing a Mann-Whitney U test comparing the observed classifier accuracy to the accuracy obtained when randomly permuting the labels of the trial types.

To assess decoding accuracy for subpopulations of neurons (mixed population, visual, auditory, AV delay-selective and delay-nonselective neurons), linear classifiers were trained with 200 randomly selected neurons belonging to these subpopulations. Overall decoding accuracy, as well as misclassification rate, was estimated for 20 iterations of the classifier. The misclassification rate was calculated as the mean false negative error for the specific misclassification between AV delay trials and auditory-shifted trials. For AV delay-selective and delay-nonselective multisensory neurons, a third classifier was trained to discriminate solely between AV delay trials. From this, the misclassification error (ms) was obtained by calculating the difference between the observed and predicted trial labels.

To characterise the decoding accuracy in different SC anatomical axes (ML, AP, DV), we varied the number of neurons used in the linear classifier. A subset of 25, 50, 75, or 100 neurons was randomly selected, and its accuracy was estimated over 20 iterations for this analysis. The final accuracy was calculated as the mean of these 20 repetitions. We also varied spatial bin size ranging from 250 µm to 650 µm in 50 µm steps. Decoding accuracy was also assessed by combining the anatomical axes. When the ML and AP axes were considered, the SC was divided into four anatomical bins (each 2.20 mm^3, 1039 neurons, s.d. ± 121). When considering all three anatomical axes (ML, AP, DV), the SC was divided into eight anatomical bins (each 1.09 mm^3, 528 neurons, s.d. ± 243). Finally, analysis on individual recordings accuracy was performed by randomly sampling 25 neurons from the available population, and the recording mean anatomical location was used. To analyse the relationship between classification performance and anatomical binning, a linear mixed model (LMM) was used. LMM analysis was used to investigate the relationship between classification accuracy along anatomical axes and the properties of the neurons used to train the decoders. This analysis utilised two sets of predictor variables. The first set comprised the proportions of visual, auditory, multisensory, and delay neurons in the training population. The second set consisted of neuron-specific properties: MII, absolute MII, FR modulation index, and reliability index. For each classification run, these variables were calculated for the neurons used to train the decoder. The distributions of these properties were then analysed, with kurtosis applied to evaluate their shape. If a distribution's kurtosis was below 2, it was considered symmetrical, and the mean was calculated. For distributions with a kurtosis above 2, indicating skewness, the median was used instead.

To assess the contribution of nonlinear interactions to decoding accuracy, we compared the performance of decoders trained on observed responses with those trained on synthetic linear sums. Both a linear classifier (SVM) and a nonlinear classifier (RFC) were used for this analysis.

Connectivity analysis

Cross-correlation analysis was performed on spiking activity between 1 s to 2 s after stimulus onset. The neuronal responses during this time window were concatenated across repetitions. For this analysis, only neurons that showed responsiveness to any stimulus presentation and with a baseline FR of at least 10 Hz were included. This resulted in 3543 neurons being included, from 92 recordings and 24 mice. The cross-correlogram (CCG) for a pair of neurons simultaneously recorded was calculated as described by Sielge et al.[39]:

$$CCG(\tau) = \frac{\frac{1}{M}\sum_{i=1}^{M}\sum_{t=1}^{N} x_1^i(t)x_2^i(t+\tau)}{\theta(\tau)\sqrt{\lambda_1\lambda_2}} \qquad \text{Eq. 6}$$

Where M is the number of trials, N is the number of bins per trial, and τ is the time lag. The spike trains of the two neurons in trial *i* are denoted by x_1^i and x_2^i. The triangular function $\theta(\tau)$ corrects for the variation in the number of overlapping bins at different time lags, while λ_1, λ_2 are the mean FR of the two neurons, used to account for FR dependence. To correct for stimulus-locked correlations and slow

temporal correlation, a jitter correction was applied, by subtracting to the original CCG a jittered cross-correlogram. A jitter window of 10 ms was used.

$$CCG_{corrected} = CCG_{original} - CCG_{jittered} \qquad \text{Eq. 7}$$

CCGs were considered significant if the peak occurred within a ± 5 ms window and exceeded five standard deviations above the mean of the noise distribution. The noise distribution was defined using the CCG values between 10–15 ms away from zero time. For each significant CCG, the peak lag was identified as the time difference between zero and the peak time. Significant correlations with a peak lag of 0 ms were excluded, as synchronous firing at 0 ms is typically attributed to shared inputs[102] rather than direct interactions. Therefore, only correlations with a peak lag between 1 ms and 5 ms were considered significant putative connections and included in further analyses.

The false positive rate of connections was estimated by performing CCG analysis on neurons sampled from different recordings, with the number of neurons considered matching those in each original recording. The depth distribution of the randomly selected neurons was also matched to that of the original neurons. This was achieved by dividing the SC depth into 13 equally spaced bins of 150 µm and selecting neurons according to these depth bins.

Sound-induced behavioural response analysis

For analysis on sound-induced behavioural responses, facial principal components were extracted using singular value decomposition (SVD) as implemented in the Facemap toolbox[103] (https://github.com/MouseLand/facemap). The average z-scored motion energy weights for the first principal component (PC) were compared across trial conditions for different animals. For decoding analysis, linear classifiers (SVM) were trained using either the first PC or the first 100 PCs to discriminate between different trial types.

Statistical analysis

To test for statistical significance between groups where appropriate we used either paired or non-paired non-parametric tests (Wilcoxon signed-ranks test or Mann–Whitney U test). To assess delay tuning of neurons we used permutation test. To compare distributions, we used the Kolmogorov–Smirnov test. Statistical test details and p values are provided in figures and/or legends.

Data availability

The data that supports the findings described in this paper will be made available upon publication. The open-source designs of the speakers and LED device and its associated circuits is available through our institute scientific hardware platform (https://github.com/FrancisCrickInstitute/Coliseum).

Code availability

Matlab and python code for analysing the data and generating the figures will be made available upon publication.

Acknowledgements

We would like to thank Tom P.A. Warner, Cecilia Della Casa, Andreas Schaefer and members of the Crick Neurophysiology community for useful discussions and feedback on the work, and Alexander Egea Weiss, Daniel R. Ward and Sadra Sadeh for comments on early versions of the manuscript. The work was supported by the Engineering and Physical Sciences Research Council (BBSRC, award ref. EP/X020924/1) and the Francis Crick Institute which receives its core funding from Cancer Research UK (10746, CC2118), the UK Medical Research Council (10746, CC2118), and the Wellcome Trust (10746, CC2118).

Contributions

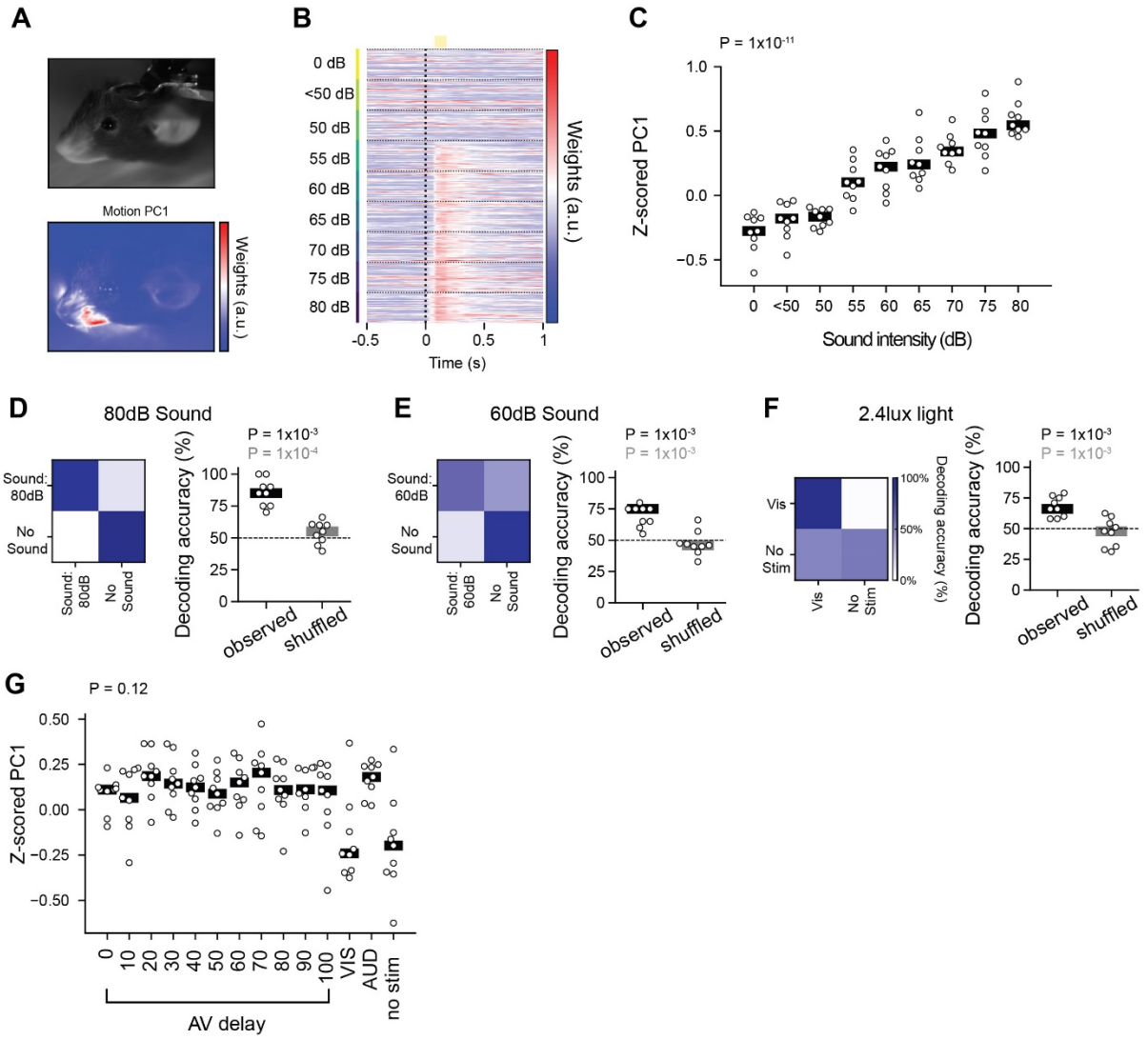
M.F.I. & G.B. conceptualized the project. G.B. curated the data, carried out formal analysis, conducted the investigation G.K., X.C.F. & A.I. developed the stimulation apparatus. I.R. conducted & designed

pilot experiments with the supervision of M.J.M. M.F.I. secured funding & provided supervision. M.F.I. & G.B. wrote the original draft. All authors reviewed and edited the manuscript.

Competing interests

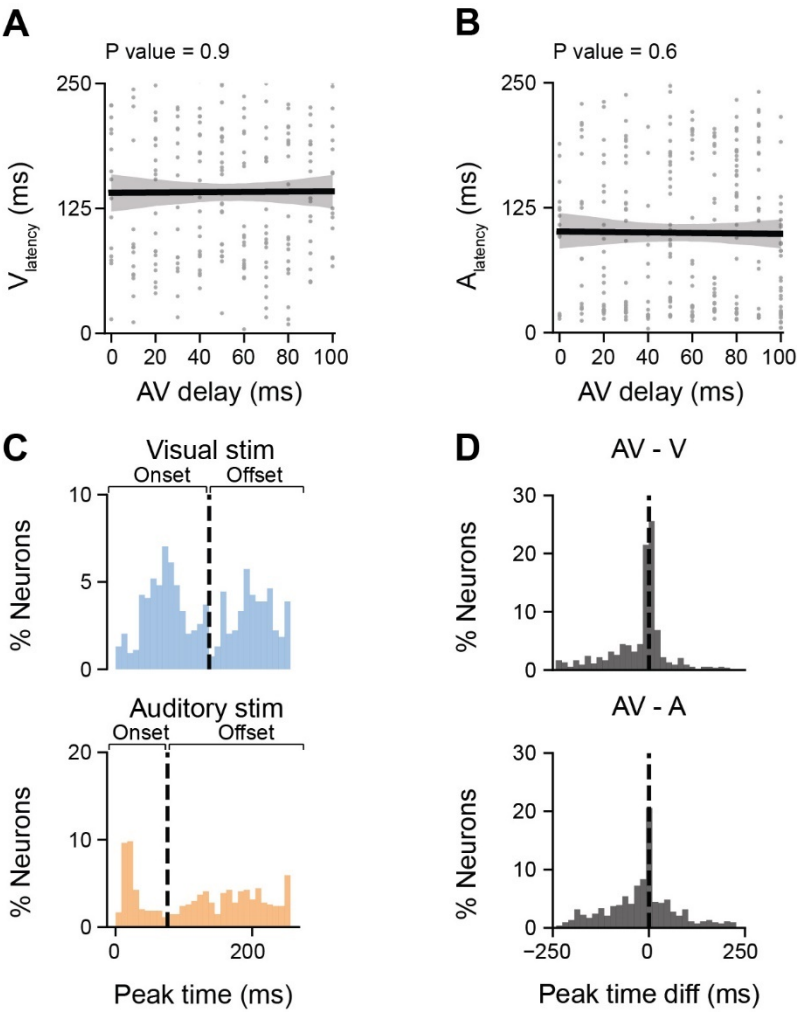
The authors declare no competing interests.

Supplementary Figures



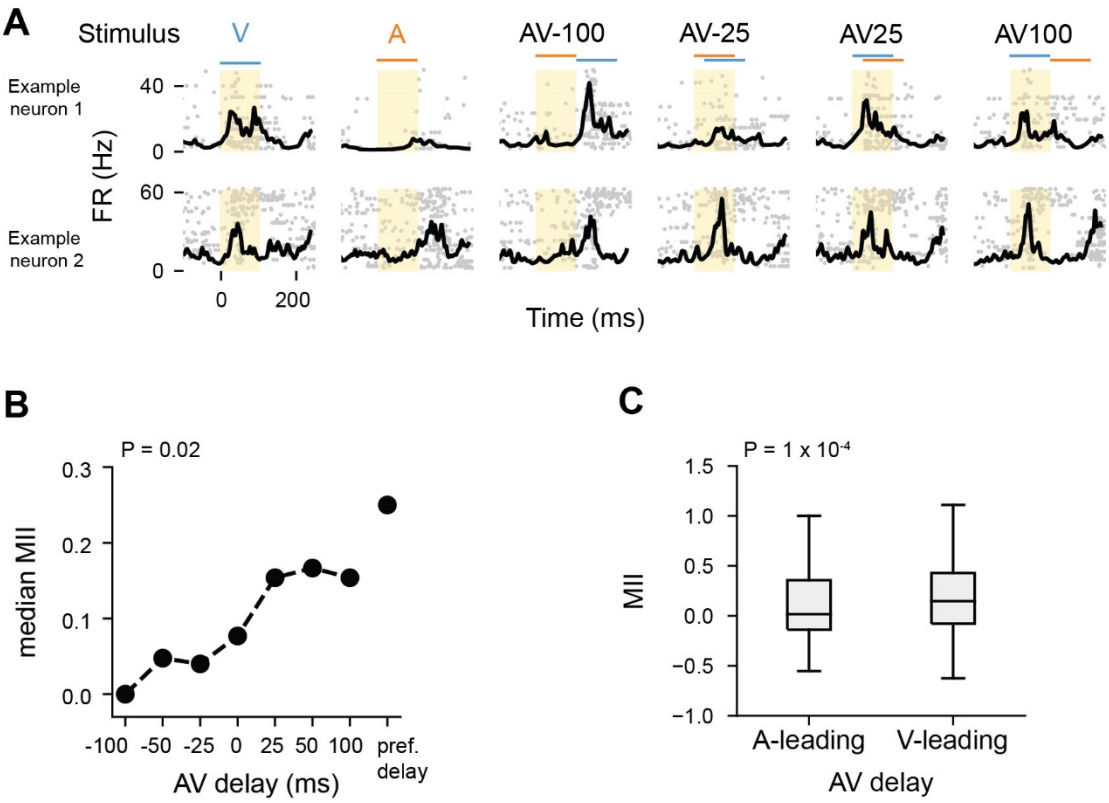
Supplementary Figure 1. Decoding sensory stimuli and audiovisual timing from motion energy representations.

A. Example video frame and spatial mask of the first principal component (PC) of the motion energy. Colour scale indicates low- (blue) to high-motion energy (red). **B**. Example representation of motion energy weights for the first PC over time. Data are aligned to stimulus onset (dashed line) and categorised by sound intensity (dB). **C**. Average z-scored motion energy weights for the first PC for the different sound intensity trials. P value from Kruskal-Wallis test. Individual datapoints represent individual recordings. **D**. Left: confusion matrix showing classification performance for sound (80 dB) versus no sound condition for an example recording. Right: decoding accuracy for classification of 80 dB sound events and no sound condition, comparing observed versus shuffled data across recordings. Black P value corresponds to Wilcoxon signed-rank test against chance value. Gray P value corresponds to Mann-Whitney U test with respect to shuffled labels. **E**. As in **D**, for classification of 60 dB sound events and no sound condition. **F**. As in **D**, for classification of visual stimulus versus no stimulus condition. **G**. Average z-scored motion energy weights for the first PC for the different trial types. P value from Kruskal-Wallis test. Individual datapoints represent individual recordings. N = 9 recordings, 7 animals. **C.** As in **B**, for a decoder trained on the first 100 principal components from the mouse videos.



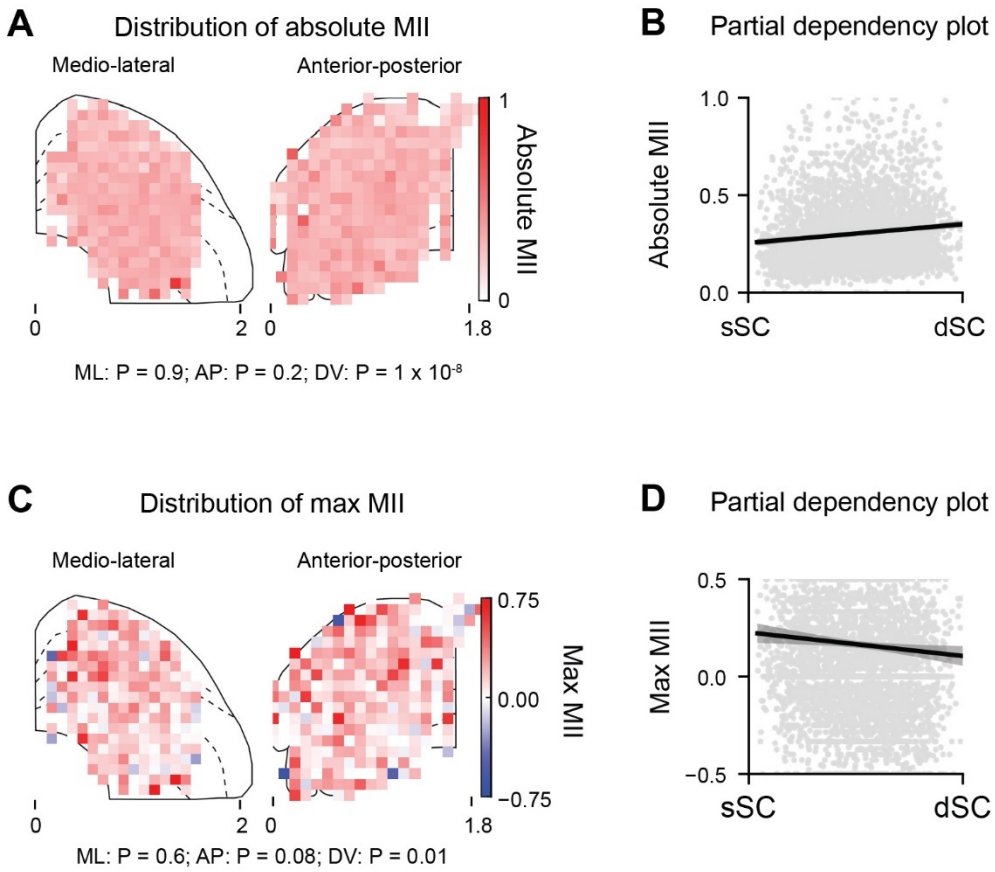
Supplementary Figure 2. Audiovisual delay is not explained by response latency to unisensory stimulation.

A. Preferred AV delay as a function of response latency to visual stimulation. Only delay-selective neurons that were responsive during visual stimulation were included (N = 232 neurons). P value from linear regression. **B**. Same as **A**, for auditory response latency. Only delay-selective neurons that were responsive during auditory stimulation were included (N = 293 neurons) **C**. Distribution of response latency to visual (blue) and auditory (orange) stimuli for bimodal neurons. **D**. Distribution of differences in peak response latency for bimodal neurons. Top: difference between AV and visual peak response latency; Bottom: difference between AV and auditory peak response latency. N = 92 recordings, 24 animals.



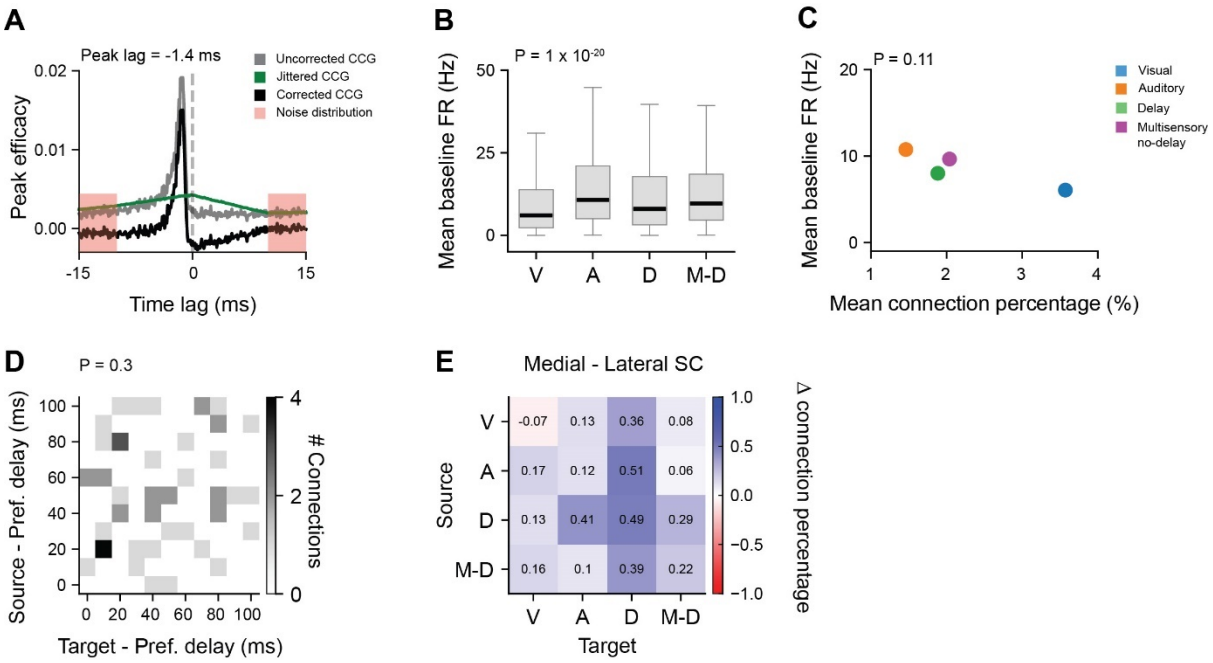
Supplementary Figure 3. Multisensory interactions are enhanced when visual stimulus leads on auditory.

A. Raster plots and PSTHs for two example neurons to unisensory and multisensory stimuli presented at variable delays. **B**. Quantification of median MII for each AV delay presented, as well as for the preferred delay across all delay neurons. P value from Kruskal-Wallis test. **C**. MII during auditory-leading stimuli and visual-leading stimuli. P value from Mann-Whitney U test. (N = 22number of neurons, recordings, 6 and animals.).



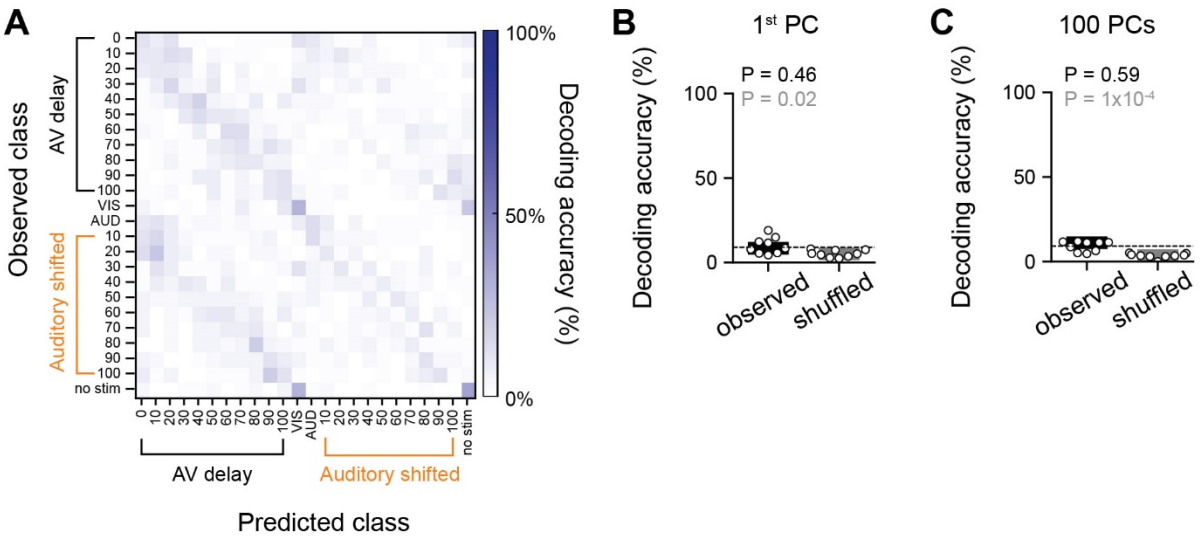
Supplementary Figure 4. Spatial distribution of neurons' multisensory index.

A. Mean absolute MII as a function of anatomical location. Left: ML axis (bins size: 102 µm ML x 120 µm DV). Centre: AP axis (bins size: 89 µm AP x 120 µm DV). Right: Partial dependency plot illustrating the linear gradient along the DV axis. P values were calculated using a linear mixed model with mouse ID as random effect. **B.** Same as A but for the maximum MII as a function of anatomical location. N = 92 recordings, 24 animals.



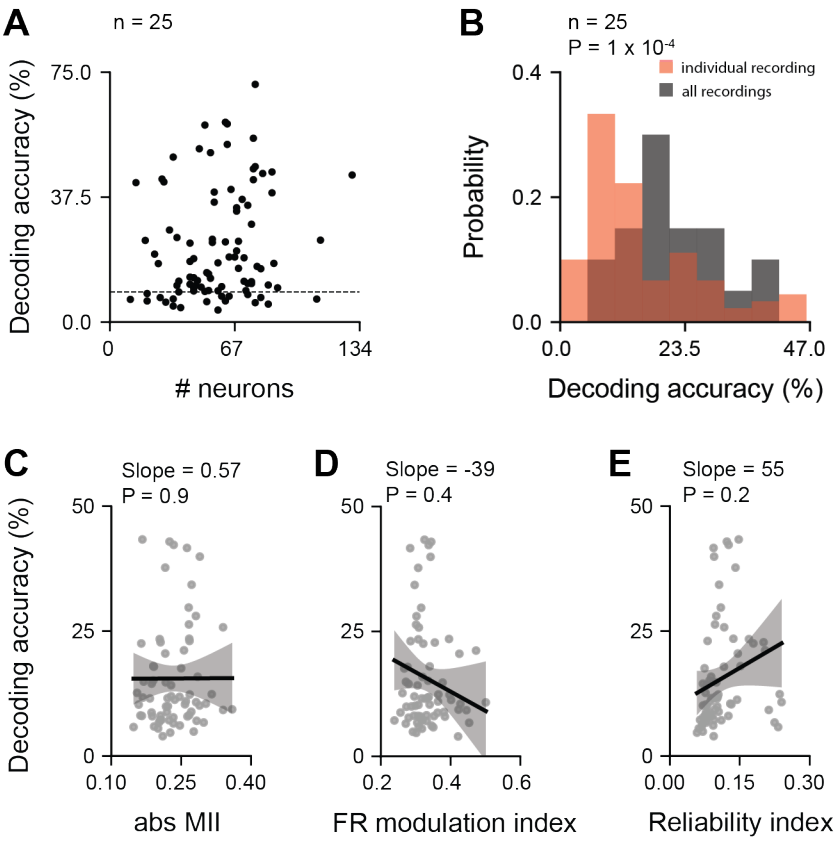
Supplementary Figure 5. Baseline firing rate of different functional subpopulations does not explain connection probabilities.

A. Jittered CCG based on all possible jitters within a 10-ms window and a noise distribution, which was estimated from the CCG flanks (10 to 15 ms before and after zero). The jittered CCG was subtracted from the original CCG to correct for slow timescale correlations larger than the jitter window[104,105]. A jitter-corrected CCG was considered significant if its peak occurred within 5 ms of zero and was more than five standard deviations above the mean of the noise distribution. **B**. Distribution of mean baseline FR for different functional subpopulations: V: visual, A: auditory, D: delay-selective neurons, M-D: multisensory delay-nonselective neurons. P value from Kruskal-Wallis test. **C**. Relationship between mean connection percentage and mean baseline FR for different functional subpopulations. P value from linear regression. **D**. Number of connections between source neuron and target neurons, binned according to their preferred delay. **E.** Matrix showing the difference in connection percentage between medial and lateral SC (bin size = 1.02 mm), for each neuronal subpopulation. Positive values (blue) indicate larger connectivity within a given class in the medial SC, whereas negative values (red) indicate larger connectivity in the lateral SC. N = 92 recordings, 24 animals.



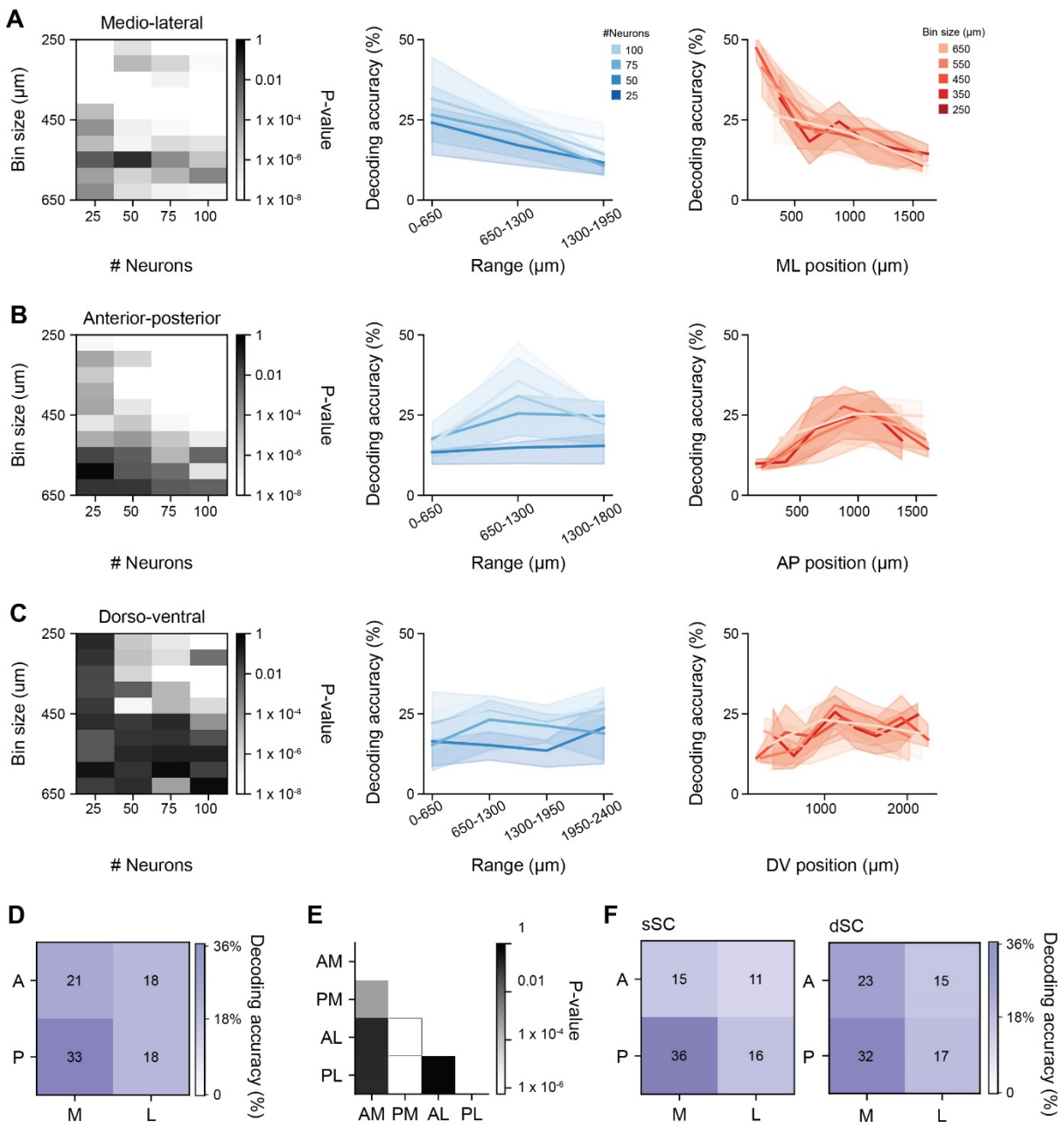
Supplementary Figure 6. Decoding audio-visual delays from motion energy representations.

A. Confusion matrix for AV delays discrimination using a decoder trained with the first PC for an example recording. **B**. Decoding accuracy for discrimination between AV delay conditions using a decoder trained with the first PC. Individual datapoints represent individual recordings. Black P value corresponds to Wilcoxon signed-rank test against chance value. Gray P value corresponds to Mann-Whitney U test with respect to shuffled labels. N = 9 recordings, 7 animals. **C**. As in **B**, for a decoder trained on the first 100 principal components from the mouse videos.



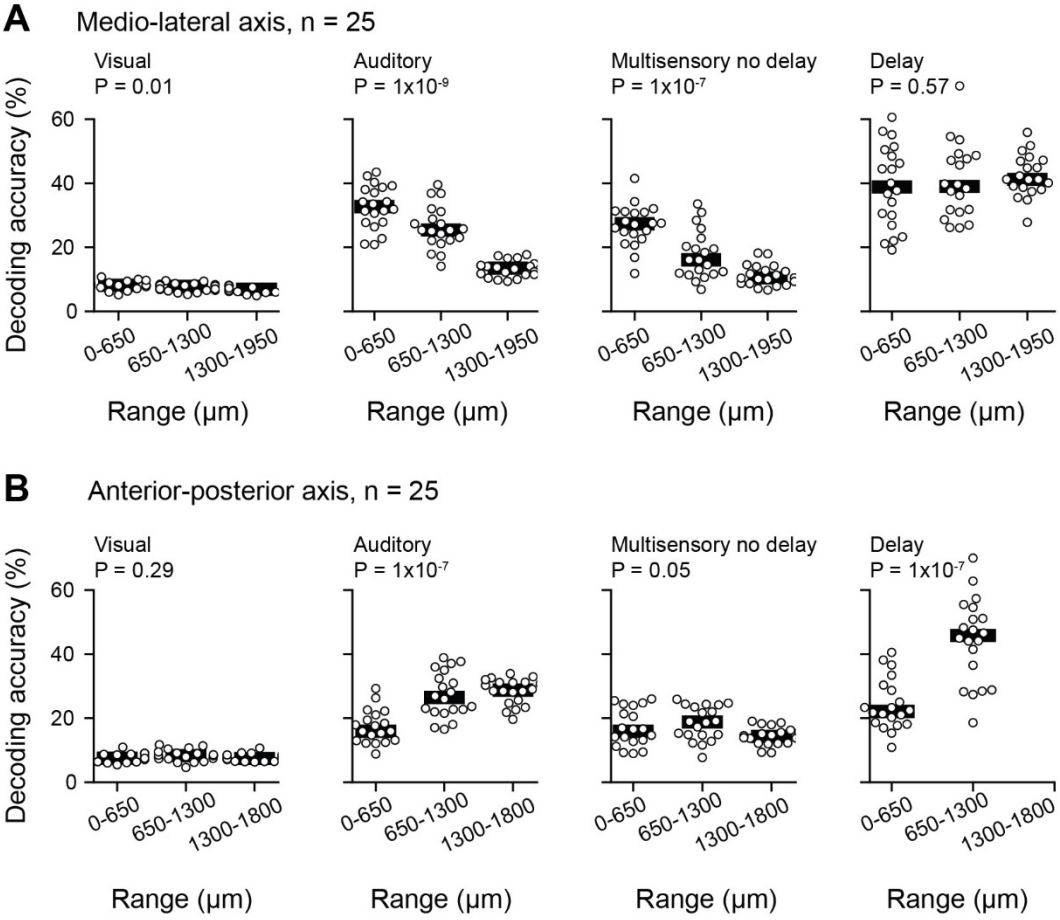
Supplementary Figure 7. The relationship between audiovisual delay decoding accuracy from individual recordings and single neuron response properties.

A. Mean decoding accuracy as a function of the number of neurons for each recording. Each circle represents an individual recording (N = 92). The dotted line represents chance level. **B**. AV delay decoding accuracy obtained from simultaneously recorded neurons (red) and randomly selected neurons from different recordings (black). Number of neurons was fixed at 25. P value from Mann-Whitney U test. **C**. Mean classifier decoding accuracy as a function of absolute MII. The black line represents the fitted regression line, while the shaded gray area denotes the 95% confidence interval. Gray circles correspond to individual recordings. P value from linear mixed model. **D**. as in **C**, for the FR modulation index. **E**. Same as **C**, for the reliability index. N = 92 recordings, 24 animals.



Supplementary Figure 8. Audiovisual delay decoding accuracy varies along the different anatomical axes of the SC.

A. AV delay decoding accuracy along the ML axis. Left: P values from Kruskal Wallis test on the mean AV delay decoding accuracy for different bin sizes and number of neurons included in the decoder. Centre: mean AV delay decoding accuracy for 25,50,75 and 100 neurons sampled from bins spaced by 650 µm along the ML axis. Right: mean AV delay decoding accuracy for 50 neurons for different bin sizes (from 250 to 650 µm). **B**. As in **A**, for the AP axis. **C**. As in **A**, for the DV axis. **D**. Mean decoding accuracy for 50 randomly sampled neurons across 4 different subregions of the SC. Each bin covered 2.20 mm^3, with 1039 ± 121 neurons per bin. **E**. Dunn's post-hoc test corrected P-values for differences in classifier accuracy among the 4 subregions shown in panel **D**. AM: anterior-medial; PM: posterior-medial; AL: anterior-lateral; PL: posterior-lateral. **F**. As in D, for 50 randomly sampled neurons across 8 different anatomical subregions of the SC. Each bin covered 1.09 mm^3, with on average 528 ± 243 neurons per bin. Left: superficial SC (sSC); Right: deep SC (dSC). N = 92 recordings, 24 animals.



Supplementary Figure 9. Contribution of functional subpopulation to gradient of AV delay encoding in the SC.

A. AV delay decoding accuracy for different functional subpopulation as a function of ML axes. Each decoder was trained on the activity of 25 randomly sampled neurons, and the procedure was repeated 20 times. **B.** As in **A** as a function of AP axes. N = 92 recordings, 24 animals.

References

1. Vroomen, J. & Keetels, M. Perception of intersensory synchrony: A tutorial review. *Atten Percept Psychophys* **72**, 871–884 (2010).

2. Alais, D., Newell, F. N. & Mamassian, P. Multisensory Processing in Review: from Physiology to Behaviour. *Seeing Perceiving* **23**, 3–38 (2010).

3. Lewis, R. & Noppeney, U. Audiovisual Synchrony Improves Motion Discrimination via Enhanced Connectivity between Early Visual and Auditory Areas. *Journal of Neuroscience* **30**, 12329–12339 (2010).

4. Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W. & Theeuwes, J. Pip and pop: nonspatial auditory signals improve spatial visual search. *J Exp Psychol Hum Percept Perform* **34**, 1053–1065 (2008).

5. McGrath, M. & Summerfield, Q. Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *J Acoust Soc Am* **77**, 678–685 (1985).

6. Shipley, T. Auditory flutter-driving of visual flicker. *Science (1979)* **145**, 1328–1330 (1964).

7. Chen, Y. C. & Spence, C. The crossmodal facilitation of visual object representations by sound: evidence from the backward masking paradigm. *J Exp Psychol Hum Percept Perform* **37**, 1784–1802 (2011).

8. Alais, D. & Burr, D. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* **14**, 257–262 (2004).

9. Lewald, J. & Guski, R. Auditory-visual temporal integration as a function of distance: No compensation for sound-transmission time in human perception. *Neurosci Lett* **357**, 119–122 (2004).

10. Jaekl, P., Seidlitz, J., Harris, L. R. & Tadin, D. Audiovisual Delay as a Novel Cue to Visual Distance. *PLoS One* **10**, e0141125 (2015).

11. Senkowski, D. & Engel, A. K. Multi-timescale neural dynamics for multisensory integration. *Nature Reviews Neuroscience 2024 25:9* **25**, 625–642 (2024).

12. Masullo, L. *et al.* Genetically Defined Functional Modules for Spatial Orienting in the Mouse Superior Colliculus. *Current Biology* **29**, 2892-2904.e8 (2019).

13. Liang, F. *et al.* Sensory Cortical Control of a Visually Induced Arrest Behavior via Corticotectal Projections. *Neuron* **86**, 755–767 (2015).

14. Hoy, J. L., Bishop, H. I. & Niell, C. M. Defined Cell Types in Superior Colliculus Make Distinct Contributions to Prey Capture Behavior in the Mouse. *Current Biology* **29**, 4130-4138.e5 (2019).

15. Wei, P. *et al.* Processing of visually evoked innate fear by a non-canonical thalamic pathway. *Nature Communications 2015 6:1* **6**, 1–13 (2015).

16. Evans, D. A. *et al.* A synaptic threshold mechanism for computing escape decisions. *Nature 2018 558:7711* **558**, 590–594 (2018).

17. Shang, C. *et al.* A parvalbumin-positive excitatory visual pathway to trigger fear responses in mice. *Science (1979)* **348**, 1472–1477 (2015).

18. Stein, B. E., Magalhães-Castro, B. & Kruger, L. Superior Colliculus: Visuotopic-Somatotopic Overlap. *Science (1979)* **189**, 224–226 (1975).

19. King, A. J. The superior colliculus. *Current Biology* **14**, R335–R338 (2004).

20. King, A. J. & Palmer, A. R. Cells responsive to free-field auditory stimuli in guinea-pig superior colliculus: distribution and response properties. *J Physiol* **342**, 361–381 (1983).

21. Meredith, M. A. & Stein, B. E. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *https://doi.org/10.1152/jn.1986.56.3.640* **56**, 640–662 (1986).

22. Ito, S., Si, Y., Litke, A. M. & Feldheim, D. A. Nonlinear visuoauditory integration in the mouse superior colliculus. *PLoS Comput Biol* **17**, e1009181 (2021).

23. Miller, R. L., Pluta, S. R., Stein, B. E. & Rowland, B. A. Relative Unisensory Strength and Timing Predict Their Multisensory Product. *Journal of Neuroscience* **35**, 5213–5220 (2015).

24. Meredith, M. A., Nemitz, J. W. & Stein, B. E. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience* **7**, 3215–3229 (1987).

25. Reinhard, K. *et al.* A projection specific logic to sampling visual inputs in mouse superior colliculus. *Elife* **8**, (2019).

26. Westby, G. W. M., Keay, K. A., Redgrave, P., Dean, P. & Bannister, M. Output pathways from the rat superior colliculus mediating approach and avoidance have different sensory properties. *Exp Brain Res* **81**, 626–638 (1990).

27. Comoli, E. *et al.* Segregated anatomical input to sub-regions of the rodent superior colliculus associated with approach and defense. *Front Neuroanat* **6**, 22244 (2012).

28. Benavidez, N. L. *et al.* Organization of the inputs and outputs of the mouse superior colliculus. *Nature Communications 2021 12:1* **12**, 1–20 (2021).

29. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature 2017 551:7679* **551**, 232–236 (2017).

30. Hafed, Z. M. & Chen, C. Y. Sharper, Stronger, Faster Upper Visual Field Representation in Primate Superior Colliculus. *Curr Biol* **26**, 1647–1658 (2016).

31. de Malmazet, D., Kühn, N. K. & Farrow, K. Retinotopic Separation of Nasal and Temporal Motion Selectivity in the Mouse Superior Colliculus. *Curr Biol* **28**, 2961-2969.e4 (2018).

32. Clayton, K. K. *et al.* Sound elicits stereotyped facial movements that provide a sensitive index of hearing abilities in mice. *Curr Biol* **34**, 1605-1620.e5 (2024).

33. Bimbard, C. *et al.* Behavioral origin of sound-evoked activity in mouse visual cortex. *Nature Neuroscience 2023 26:2* **26**, 251–258 (2023).

34. Drager, U. C. & Hubel, D. H. Physiology of visual cells in mouse superior colliculus and correlation with somatosensory and auditory input. *Nature* **253**, 203–204 (1975).

35. Montijn, J. S. *et al.* A parameter-free statistical test for neuronal responsiveness. *Elife* **10**, (2021).

36. 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).

37. Jay, M. F. & Sparks, D. L. Sensorimotor integration in the primate superior colliculus. II. Coordinates of auditory signals. *J Neurophysiol* **57**, 35–55 (1987).

38. Trepka, E. B., Zhu, S., Xia, R., Chen, X. & Moore, T. Functional interactions among neurons within single columns of macaque V1. *Elife* **11**, (2022).

39. Siegle, J. H. *et al.* Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature 2021 592:7852* **592**, 86–92 (2021).

40. Gehr, C., Sibille, J. & Kremkow, J. Retinal input integration in excitatory and inhibitory neurons in the mouse superior colliculus in vivo. *Elife* **12**, (2023).

41. Perkel, D. H., Gerstein, G. L. & Moore, G. P. Neuronal spike trains and stochastic point processes. I. The single spike train. *Biophys J* **7**, 391–418 (1967).

42. Moore, G. P., Segundo, J. P., Perkel, D. H. & Levitan, H. Statistical signs of synaptic interaction in neurons. *Biophys J* **10**, 876–900 (1970).

43. Narayanan, N. S., Kimchi, E. Y. & Laubach, M. Redundancy and Synergy of Neuronal Ensembles in Motor Cortex. *Journal of Neuroscience* **25**, 4207–4216 (2005).

44. Hazon, O. *et al.* Noise correlations in neural ensemble activity limit the accuracy of hippocampal spatial representations. *Nature Communications 2022 13:1* **13**, 1–13 (2022).

45. Dwulet, J. M. *et al.* Correlated spontaneous activity sets up multi-sensory integration in the developing higher-order cortex. *bioRxiv* 2024.07.19.603239 (2024) doi:10.1101/2024.07.19.603239.

46. Bertini, C., Leo, F. & Làdavas, E. Temporo-nasal asymmetry in multisensory integration mediated by the Superior Colliculus. *Brain Res* **1242**, 37–44 (2008).

47. Rowland, B. A., Stanford, T. R. & Stein, B. E. A Model of the Neural Mechanisms Underlying Multisensory Integration in the Superior Colliculus. *http://dx.doi.org/10.1068/p5842* **36**, 1431–1443 (2007).

48. Rowland, B. A. & Stein, B. E. Temporal profiles of response enhancement in multisensory integration. *Front Neurosci* **2**, 218–224 (2008).

49. Rowland, B. A. & Stein, B. E. A model of the temporal dynamics of multisensory enhancement. *Neurosci Biobehav Rev* **41**, 78–84 (2014).

50. Probing the integration of auditory, visual and motor signals in mouse superior colliculus. https://www.abstractsonline.com/pp8/?_gl=1*gghpl7*_gcl_au*OTM2MDI0NzQ2LjE3MzA1NTEzMzI.*_ga*MTc1MzQ1Nzc2OS4xNzMwNTUxMzMy*_ga_T09K3Q2WDN*MTczMDU1MTMzMi4xLjEuMTczMDU1MTMzMi42MC4wLjA.#!/10619/presentation/83309.

51. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. *Curr Opin Neurobiol* **37**, 66–74 (2016).

52. Jeffrey Johnston, W., Palmer, S. E. & Freedman, D. J. Nonlinear mixed selectivity supports reliable neural computation. *PLoS Comput Biol* **16**, e1007544 (2020).

53. May, P. J. The mammalian superior colliculus: laminar structure and connections. *Prog Brain Res* **151**, 321–378 (2006).

54. Li, C. *et al.* Pathway-specific inputs to the superior colliculus support flexible responses to visual threat. *Sci Adv* **9**, (2023).

55. Alvarado, J. C., Rowland, B. A., Stanford, T. R. & Stein, B. E. A neural network model of multisensory integration also accounts for unisensory integration in superior colliculus. *Brain Res* **1242**, 13–23 (2008).

56. Cuppini, C., Stein, B. E. & Rowland, B. A. Development of the Mechanisms Governing Midbrain Multisensory Integration. *Journal of Neuroscience* **38**, 3453–3465 (2018).

57. Alvarado, J. C., Stanford, T. R., Vaughan, J. W. & Stein, B. E. Cortex Mediates Multisensory But Not Unisensory Integration in Superior Colliculus. *Journal of Neuroscience* **27**, 12775–12786 (2007).

58. Binns, K. E. & Salt, T. E. Importance of NMDA receptors for multimodal integration in the deep layers of the cat superior colliculus. *https://doi.org/10.1152/jn.1996.75.2.920* **75**, 920–930 (1996).

59. Behabadi, B. F., Polsky, A., Jadi, M., Schiller, J. & Mel, B. W. Location-Dependent Excitatory Synaptic Interactions in Pyramidal Neuron Dendrites. *PLoS Comput Biol* **8**, e1002599 (2012).

60. Major, G., Polsky, A., Denk, W., Schiller, J. & Tank, D. W. Spatiotemporally graded NMDA spike/plateau potentials in basal dendrites of neocortical pyramidal neurons. *J Neurophysiol* **99**, 2584–2601 (2008).

61. Branco, T., Clark, B. A. & Häusser, M. Dendritic discrimination of temporal input sequences in cortical neurons. *Science (1979)* **329**, 1671–1675 (2010).

62. Kardamakis, A. A., Pérez-Fernández, J. & Grillner, S. Spatiotemporal interplay between multisensory excitation and recruited inhibition in the lamprey optic tectum. *Elife* **5**, (2016).

63. Ursino, M., Cuppini, C., Magosso, E., Serino, A. & Pellegrino, G. Multisensory integration in the superior colliculus: A neural network model. *J Comput Neurosci* **26**, 55–73 (2009).

64. Ohshiro, T., Angelaki, D. E. & Deangelis, G. C. A normalization model of multisensory integration. *Nature Neuroscience 2011 14:6* **14**, 775–782 (2011).

65. Scholl, B., Wilson, D. E. & Fitzpatrick, D. Local Order within Global Disorder: Synaptic Architecture of Visual Space. *Neuron* **96**, 1127-1138.e4 (2017).

66. Chen, T. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature 2013 499:7458* **499**, 295–300 (2013).

67. Dana, H. *et al.* High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nature Methods 2019 16:7* **16**, 649–657 (2019).

68. Iacaruso, M. F., Gasler, I. T. & Hofer, S. B. Synaptic organization of visual space in primary visual cortex. *Nature 2017 547:7664* **547**, 449–452 (2017).

69. Fuentes-Santamaria, V., Alvarado, J. C., McHaffie, J. G. & Stein, B. E. Axon Morphologies and Convergence Patterns of Projections from Different Sensory-Specific Cortices of the Anterior Ectosylvian Sulcus onto Multisensory Neurons in the Cat Superior Colliculus. *Cerebral Cortex* **19**, 2902–2915 (2009).

70. Stein, B. E., Meredith, M. A. & Wallace, M. T. The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Prog Brain Res* **95**, 79–90 (1993).

71. Gouras, P. The effects of light-adaptation on rod and cone receptive field organization of monkey ganglion cells. *J Physiol* **192**, 747–760 (1967).

72. Neely, S. T., Norton, S. J., Gorga, M. P. & Jesteadt, W. Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli. *J Acoust Soc Am* **83**, 652–656 (1988).

73. Zampini, M. & Spence, C. Assessing the Role of Visual and Auditory Cues in Multisensory Perception of Flavor. *The Neural Bases of Multisensory Processes* 739–757 (2012) doi:10.1201/b11092-47.

74. Stein, B. E., Stanford, T. R. & Rowland, B. A. Development of multisensory integration from the perspective of the individual neuron. *Nat Rev Neurosci* **15**, 520 (2014).

75. Sarko, D. K. *et al.* Spatial and Temporal Features of Multisensory Processes. *The Neural Bases of Multisensory Processes* 191–215 (2012).

76. Dear, S. P. & Suga, N. Delay-tuned neurons in the midbrain of the big brown bat. *J Neurophysiol* **73**, 1084–1100 (1995).

77. Wenstrup, J. J. & Portfors, C. V. Neural processing of target distance by echolocating bats: functional roles of the auditory midbrain. *Neurosci Biobehav Rev* **35**, 2073–2083 (2011).

78. Posner, M. I. Orienting of attention. *Q J Exp Psychol* **32**, 3–25 (1980).

79. Rafal, R., Henik, A. & Smith, J. Extrageniculate contributions to reflex visual orienting in normal humans: a temporal hemifield advantage. *J Cogn Neurosci* **3**, (1991).

80. Ansorge, U. Asymmetric influences of temporally vs. nasally presented masked visual information: Evidence for collicular contributions to nonconscious priming effects. *Brain Cogn* **51**, 317–325 (2003).

81. Kristjánsson, Á., Vandenbroucke, M. W. G. & Driver, J. When pros become cons for anti- versus prosaccades: Factors with opposite or common effects on different saccade types. *Exp Brain Res* **155**, 231–244 (2004).

82. Johnson, M. H. Cortical Maturation and the Development of Visual Attention in Early Infancy. *J Cogn Neurosci* **2**, 81–95 (1990).

83. Jóhannesson, Ó. I., Tagu, J. & Kristjánsson, Á. Asymmetries of the visual system and their influence on visual performance and oculomotor dynamics. *Eur J Neurosci* **48**, 3426–3445 (2018).

84. Sherman, S. M. Visual Fields of Cats with Cortical and Tectal Lesions. *Science (1979)* **185**, 355–357 (1974).

85. Ito, S. & Feldheim, D. A. The mouse superior colliculus: An emerging model for studying circuit formation and function. *Front Neural Circuits* **12**, 327065 (2018).

86. Vetter, P., Smith, F. W. & Muckli, L. Decoding Sound and Imagery Content in Early Visual Cortex. *Current Biology* **24**, 1256–1262 (2014).

87. Vetter, P. *et al.* Decoding Natural Sounds in Early 'Visual' Cortex of Congenitally Blind Individuals. *Curr Biol* **30**, 3039-3044.e2 (2020).

88. Wang, M. *et al.* Retinotopic organization of feedback projections in primate early visual cortex: implications for active vision. *bioRxiv* 2022.04.27.489651 (2022) doi:10.1101/2022.04.27.489651.

89. Wang, L., Liu, M., Segraves, M. A. & Cang, J. Visual Experience Is Required for the Development of Eye Movement Maps in the Mouse Superior Colliculus. *Journal of Neuroscience* **35**, 12281–12286 (2015).

90. Hu, F. & Dan, Y. An inferior-superior colliculus circuit controls auditory cue-directed visual spatial attention. *Neuron* **110**, 109-119.e3 (2022).

91. Essig, J., Hunt, J. B. & Felsen, G. Inhibitory neurons in the superior colliculus mediate selection of spatially-directed movements. *Communications Biology 2021 4:1* **4**, 1–14 (2021).

92. Wang, L., McAlonan, K., Goldstein, S., Gerfen, C. R. & Krauzlis, R. J. A Causal Role for Mouse Superior Colliculus in Visual Perceptual Decision-Making. *Journal of Neuroscience* **40**, 3768–3782 (2020).

93. Cregg, J. M. *et al.* Brainstem neurons that command mammalian locomotor asymmetries. *Nature Neuroscience 2020 23:6* **23**, 730–740 (2020).

94. Steinmetz, N. A. *et al.* Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science (1979)* **372**, (2021).

95. Bouin, A. *et al.* New rabies viral resources for multi-scale neural circuit mapping. *Molecular Psychiatry 2024 29:7* **29**, 1951–1967 (2024).

96. Xu, X. *et al.* Viral vectors for neural circuit mapping and recent advances in trans-synaptic anterograde tracers. *Neuron* **107**, 1029 (2020).

97. Suter, K. J., Smith, B. N. & Dudek, F. E. Electrophysiological recording from brain slices. *Methods* **18**, 86–90 (1999).

98. Okun, M. *et al.* Diverse coupling of neurons to populations in sensory cortex. *Nature 2015 521:7553* **521**, 511–515 (2015).

99. Mianné, J. *et al.* Correction of the auditory phenotype in C57BL/6N mice via CRISPR/Cas9-mediated homology directed repair. *Genome Med* **8**, 1–12 (2016).

100. Ragan, T. *et al.* Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nat Methods* **9**, 255–258 (2012).

101. Tyson, A. L. *et al.* Accurate determination of marker location within whole-brain microscopy images. *Scientific Reports 2022 12:1* **12**, 1–8 (2022).

102. Ostojic, S., Brunel, N. & Hakim, V. How Connectivity, Background Activity, and Synaptic Properties Shape the Cross-Correlation between Spike Trains. *Journal of Neuroscience* **29**, 10234–10253 (2009).

103. Syeda, A. *et al.* Facemap: a framework for modeling neural activity based on orofacial tracking. *Nature Neuroscience 2023 27:1* **27**, 187–195 (2023).

104. Harrison, M. T. & Geman, S. A Rate and History-Preserving Resampling Algorithm for Neural Spike Trains. *Neural Comput* **21**, 1244–1258 (2009).

105. Smith, M. A. & Kohn, A. Spatial and Temporal Scales of Neuronal Correlation in Primary Visual Cortex. *Journal of Neuroscience* **28**, 12591–12603 (2008).