

Master Thesis

**Effect of early visual experience on audiovisual processing in
mouse visual cortex**

by

Matúš Halák

(2724858)

Daily supervisor: Huub Terra, PhD (h.terra@nin.knaw.nl)
VU supervisor: Huib Mansvelder, PhD (h.d.mansvelder@vu.nl)

*Submitted in fulfillment of the requirements
for the
Master of Science
in
Neurosciences
at the
Vrije Universiteit Amsterdam*

Effect of early visual experience on audiovisual processing in mouse visual cortex

Matúš Halák

Vrije Universiteit Amsterdam, Netherlands Institute of Neuroscience
Amsterdam, The Netherlands
matus.halak@gmail.com

ABSTRACT

Seamless integration of parallel sources of sensory information is one of the key mechanisms that allows biological sensory systems to efficiently form complex percepts of the environment. While multisensory integration is thought to develop through experience, the role of individual neurons across visuocortical regions in this process remains unmapped. Here, we studied the effect of early visual experience on audiovisual processing in mice reared in darkness from birth. Using two-photon calcium imaging, we characterized the isolated and paired processing of visual and auditory modalities in neurons across the visual cortex of awake mice. Regression modeling enabled us to dissociate sensory and movement-related components of single-neuron activity across visuocortical regions, and thereby isolate the effects of early visual experience on multisensory processing. Our results showed that neurons in primary visual cortex (V1) and dorsal stream higher visual areas (HVAs) require early visual experience to establish a balance between visual and auditory processing. Namely, we find that early visual deprivation is associated with a shift toward auditory and away from visual processing, as well as, impaired audiovisual integration in both V1 and dorsal stream HVAs. Together, these results point to a common, experience-dependent plasticity mechanism that calibrates visual and auditory drives in both V1 and dorsal stream. By mapping this experience-dependent plasticity across V1 and dorsal stream HVAs at cellular resolution, our study provides insights into possible circuit-level mechanisms of audiovisual integration in the visual cortex.

1 INTRODUCTION

The ability to combine incoming sensory information across multiple modalities is a hallmark of animal sensory systems [1–5]. This mode of information processing, termed multisensory integration, can be observed throughout the brain's sensory regions [2]. Multisensory integration helps animals deal with uncertainty and noise arising from both the physics of the observed world and the imperfections of biological sensory organs, by resolving several independent streams of sensory information into unified percepts that can guide cognition and action [3]. Crucially, despite the universal need of animals to synthesize information across their available senses to optimally interact with their environment, this capability is largely absent at birth [4]. Rather, multisensory integration emerges through early-life experience to optimize the processing of cross-modal associations frequently encountered in an individual's environment [6].

By now a wealth of scientific evidence supports the idea that the sensory brain regions exhibit critical and sensitive periods in early development. During these developmental stages, the newborn

neural circuitry undergoes widespread plasticity and refinement through sensory experiences in relevant real-world environments [6]. Historically, much of the early progress in characterizing critical periods in vision came from landmark studies using visual deprivation and dark rearing in newborn animals [7, 8]. This line of research showed that the most foundational levels of the visual system, responsible for detecting simple visual features, required early visual experience to develop properly. Later studies of critical periods for audiovisual integration have used additional paradigms involving rearing animals in artificial environments with random sensory cues or anomalous cross-modal associations to show that the extent and form of multisensory integration are shaped by experience during early critical periods [4].

Visual critical periods have also been extensively studied in humans suffering from congenital blindness due to dense bilateral cataracts [4]. These patients live the first months of their lives, overlapping with the visual critical periods, in complete absence of visual acuity. While typically these patients regain most visual functions following cataract surgery a few months after birth, the lack of visual experience during the critical period after birth results in some life-long impairments. Specifically, even after sight restoration congenital cataract patients struggle with voice-face recognition, audiovisual speech perception, and other multisensory functions, such as multisensory enhancement, thought to be dependent on early sensory experience [9–11]. These patients also show increased cortical thickness across the visual cortex, reflecting the months-long arrest in experience-dependent synaptic pruning, one of the essential mechanisms of circuit refinement during the visual critical period after birth [11]. Additionally, sight-restored cataract patients have elevated responses to auditory signals in visual brain areas [12]. Altogether, these results suggest that a brief period of postnatal visual deprivation is sufficient to trigger cross-modal reorganization of the visual cortex with persistent functional consequences for audiovisual processing and integration. Despite clear evidence of lasting cross-modal reorganization, how individual neurons across different areas of the visual cortex adapt to postnatal visual deprivation remains uncharacterized.

The visual cortex consists of a set of hierarchically organized areas responsible for distributed processing of visual features [13, 14]. Simple visual features are first extracted in the primary visual cortex (V1) before being processed further in higher visual areas (HVAs). Information about these visual features follows two canonical pathways; the 'dorsal stream' areas convey and process motion and action-related visual information while the 'ventral stream' areas deal with shape and object recognition and categorization. A range of human developmental disorders associated with dorsal stream deficits (e.g. motion perception) has led clinicians and scientists

to believe that the dorsal stream is ‘vulnerable’ in early development [15]. Converging evidence from rodent research indicates that dorsal stream HVA neurons exhibit a developmental delay in tuning and require visual experience to develop fully, whereas ventral HVAs are more resistant to lack of visual experience and have well-tuned visual neurons soon after eye opening [16]. Still, it remains unclear whether audiovisual integration is subject to the same dorsal stream sensitive period as visual processing.

In general, multisensory neurons are distributed widely across the sensory cortices leading some scientists to consider the entire neocortex as a multisensory structure [2]. Within the visual system, HVAs are traditionally considered to be the primary loci of multisensory integration, because of their connectivity and proximity to the primary somatosensory and primary auditory cortices. Indeed, several HVAs have been shown to enhance the detection of weak multisensory stimuli via multisensory enhancement [17, 18]. However, V1 neurons have also been shown to perform numerous multimodal functions ranging from congruency-dependent visual response modulation [19], enhancement of auditory responses [20], to sharpening orientation-selectivity [21] and improving visual stimulus encoding [22]. Moreover, multisensory neurons are confined to neither the dorsal nor ventral visual stream [23, 24]. Therefore, precisely which areas of the visual system require early visual experience to develop their multisensory functions remains an open question.

Beyond identifying candidate areas, understanding experience-dependent plasticity in audiovisual processing requires functional mapping of the visual cortex across a range of spatial scales. On a large scale, a respected hypothesis is that visual cortex neurons engaged in auditory processing concentrate around borders between visual and auditory cortices [25]. This view is supported by evidence that proportion of tone-responsive V1 neurons is higher near the lateral V1 border [19]. At the same time, recent work shows a shift toward auditory processing in dorsal stream regions of dark-reared mice [26], highlighting the possibility that the balance between visual and nonvisual neuron distribution in the visual dorsal stream might be established by experience. On a microcircuit scale, neurons with similar visual feature tuning have been shown to become preferentially connected even in complete absence of visual experience [27, 28], indicating that like-to-like connectivity might be a general organizing principle of local connectivity in the visual cortex. A study investigating multisensory integration in a visuotactile HVA extended this idea to show that unimodal neurons responding to the same modality form microclusters embedded between multimodal neurons [17]. Whether similar like-to-like local organization exists in other visual areas also for neurons involved in auditory processing, and whether this modality-specific organization requires visual experience is still unknown.

Here, we explore the effect of postnatal visual deprivation on audiovisual processing in V1 and across both ventral stream (lateral-medial - LM) and dorsal stream (anteromedial - AM; posteromedial PM; anterior - A; rostral-lateral - RL; anterolateral - AL) HVAs. We modeled visual deprivation by dark rearing mice from birth. To simultaneously achieve single-neuron, as well as population-level resolution, we performed two-photon calcium imaging while presenting moving bars and sounds to awake, head-fixed, and freely running mice. Following recent work [29–32], we investigated and

addressed the confounding influence of movements on neural responses to audiovisual stimuli. Dark rearing influenced stimulus-evoked movements, the contribution of which to neural activity was dissociable from the audiovisual stimulus contribution. Based on previous research [15, 16, 26], we hypothesized that dark rearing would selectively affect audiovisual processing in the dorsal stream, leading to enhanced responses to auditory stimuli and reduced responses to visual stimuli; accompanied by deficits in audiovisual integration. We find that the visual cortex of vision-deprived mice shows cross-modal functional reorganization not restricted to the dorsal stream. Effects of dark rearing included enhanced representation of auditory and reduced representation of visual stimuli; increased proportion of multimodal and decreased proportion of visual neurons, as well as, changes in local clustering of visual, auditory and multimodal neurons. V1 and dorsal stream neurons of dark-reared mice were further characterized by lesser responsiveness and direction selectivity to visual stimuli and impaired multisensory integration. Our study is among the first to provide single-neuron insight across V1 and HVAs on the motor influences on neural activity; as well as, microarchitectural and multisensory functional organization and how these are shaped by early visual experience. The results can contribute to understanding not only the pathology arising from lack of early visual experience, but also the organizing principles of structural and functional plasticity in the visual cortex.

2 RESULTS

To study the effect of early visual experience on audiovisual processing and multisensory integration, we passively presented combinations of a moving bar and a moving sound to naive mice (Fig. 1A). Paired presentations of moving bar together with moving sound were either congruent (bar and sound moving in same direction) or incongruent (bar and sound moving in opposite directions). We used two groups of mice, one group reared in complete darkness from birth (dark-reared - DR), and a second group reared in normal light/dark conditions (normally reared - NR). During the experiment, activity of individual L2/3 neurons was recorded using two-photon calcium imaging in awake, head-fixed, freely running mice (Fig. 1B). The audiovisual stimuli were presented in absence of background light and facial movements (whisker motion, pupil size) and run-speed were recorded under infrared light. Across 19 mice (11 NR; 8 DR) and 21 recording sessions (12 NR; 9 DR), we recorded from neurons across V1 and HVAs (NR: 2868 neurons; DR: 2264 neurons) (Fig. 1C). Because of insufficient number of recording sessions in individual dorsal stream HVAs and evidence for functional separation of the dorsal stream into medial and lateral parts [33], we grouped dorsal stream HVAs into medial dorsal stream (AM/PM) and lateral dorsal stream (A/RL/AL) regions.

2.1 Dark rearing affects stimulus-evoked movements correlated with neural activity

Considering recent evidence that movements related to running, whisking and pupil size are key drivers of cortical activity [29–31] and could thereby majorly confound our findings, we investigated stimulus-evoked movements in dark-reared and normally reared mice (Fig. 2A). We found that both dark-reared and normally reared

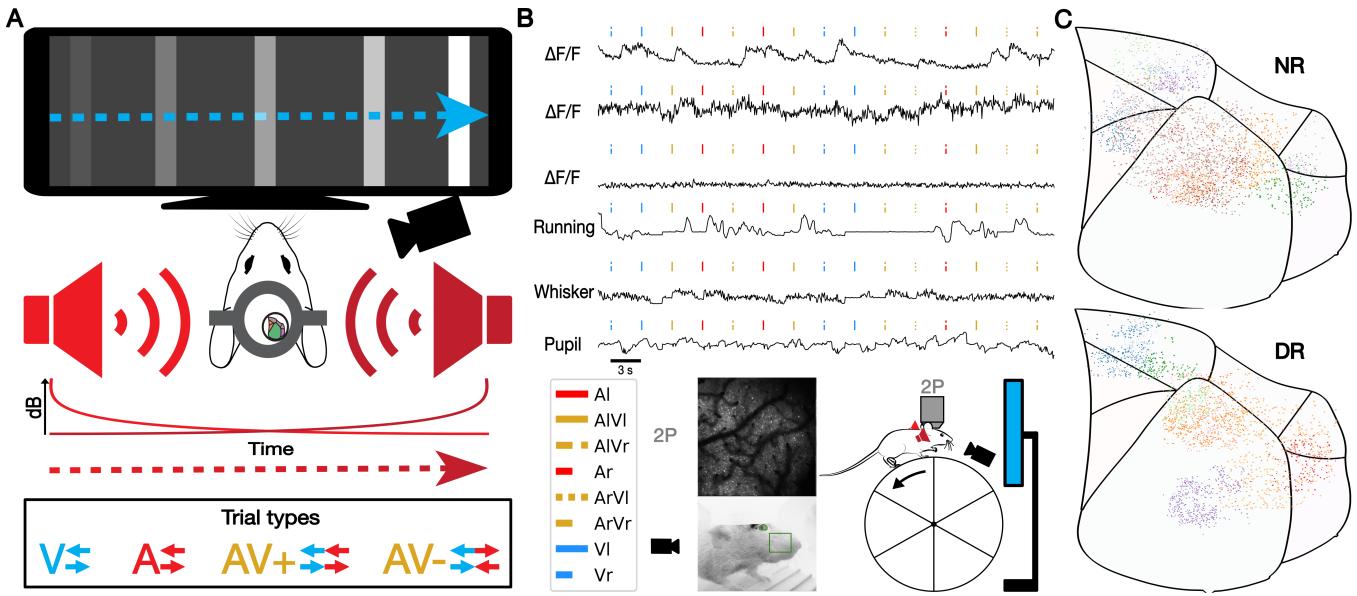


Figure 1: Two-photon calcium imaging experiment with audiovisual stimuli. **A)** Stimuli and resulting trial-types (V: visual trials, A: auditory trials, AV+: congruent audiovisual trials, AV-: incongruent audiovisual trials). **B)** Two-photon calcium imaging experimental setup (bottom right) and examples of raw recorded data (bottom middle) and extracted signals (top). **C)** All recorded neurons (NR: 2868 neurons; DR: 2264 neurons) colored based on recording session (NR: 12 sessions, 11 mice; DR: 9 sessions, 8 mice).

mice exhibited more running and whisking during auditory compared to visual trials (Fig. 2B). In both groups, auditory stimuli evoked a slight pupil dilation (sign of increased arousal [34]), while presenting high contrast moving visual stimuli led to pupil constriction (sign of pupillary contrast response [35]). Surprisingly, dark-reared mice showed more pronounced stimulus-evoked running but less pronounced stimulus-evoked whisker movement than normally reared mice, especially in auditory and audiovisual trials. Apart from differences in stimulus-evoked movements between dark-reared mice and normally reared mice, there was also variation in the amplitude and shape of stimulus-evoked movements between different normally reared (Extended Data Fig. 1A) and dark-reared animals and between experimental sessions (Extended Data Fig. 1B).

We next checked whether these stimulus-evoked movements correlated with population neural activity on a trial level (Fig. 2C). Based on previous research showing that whisker movements explain most of sound-evoked V1 activity relative to other movements [31], we expected whisking to be strongly correlated with neural activity during auditory and audiovisual trials. Instead, we found whisking to be weakly correlated with population neural activity, with no evidence of differences between different stimulus modalities. In contrast, running was moderately positively correlated with population neural activity, especially in auditory and audiovisual trials. Additionally, pupil size was weakly negatively correlated with neural activity during visual trials, but not auditory or audiovisual trials. These results indicated that our audiovisual stimuli reliably evoked bodily and facial movements correlated with neural activity, and that dark rearing affected the magnitude of the evoked movements.

2.2 Dark rearing weakens visual but strengthens auditory representations in V1 and dorsal stream neurons

Our analysis of stimulus-evoked movements raised the possibility that the effects of dark rearing on stimulus-evoked movements could account for differences in neural activity between dark-reared and normally reared mice. This was especially problematic because 75% of our trials used auditory stimuli which are known to evoke movements that explain significant portions of cortical activity [31, 32]. Nevertheless, recent work has shown that visual, auditory and motor contributions to neural activity in V1 are dissociable [32]. Thus, drawing from previous work [30, 32, 36], to isolate the sensory components of neural activity, we built a regression model (Fig. 3A, Extended Data Fig. 2A-B) for predicting single-neuron activity based on visual and auditory stimulus features, as well as motor activity (including features engineered from running, whisking and pupil size movement signals). Our model successfully dissociated the temporally-overlapping influence of sensory stimuli and motor activity on single-neuron activity across trial types (Extended Data 3A-C), although the model performed especially well in visual and audiovisual trials (Extended Data 3C).

Using this approach, we compared the contribution of visual, auditory and motor predictors to single-neuron responses across the visual cortex to visual, auditory and audiovisual stimuli. In both dark-reared and normally reared animals visual stimuli explained most variance in neural responses to both visual (Extended Data Fig. 2D) and audiovisual stimuli (Fig. 3B). Furthermore, across visuocortical regions, visual stimuli explained less variance in single-neuron activity of dark-reared mice compared to normally reared

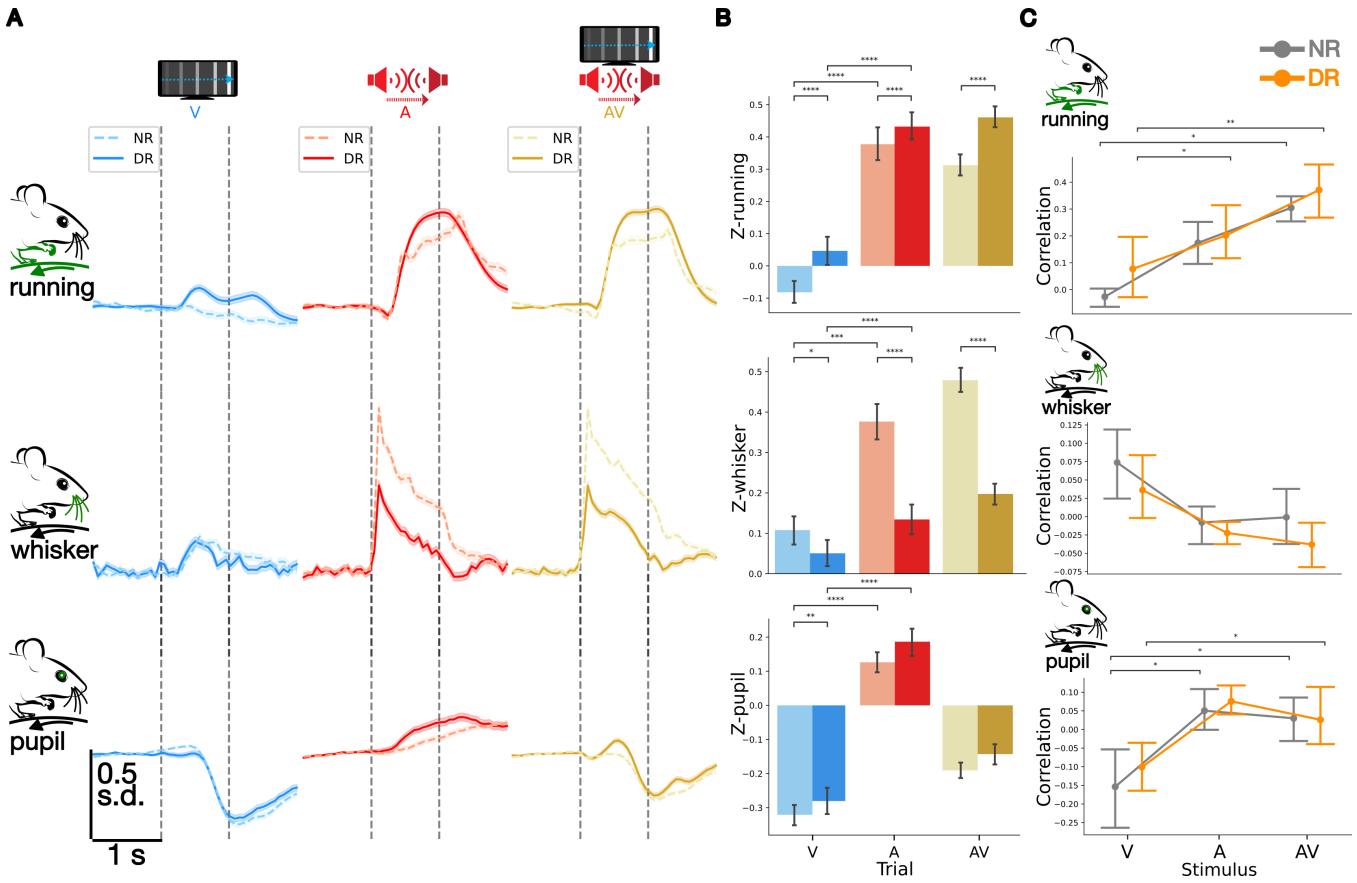


Figure 2: Dark rearing affects stimulus-evoked movements correlated with neural activity. **A)** Average movement signals across trial types and rearing conditions. Shaded regions indicate ± 1 SEM. **B)** Average stimulus-evoked movement responses across trial types and rearing conditions. **C)** Average correlation between z-scored movement signals and z-scored population neural activity across trial types and rearing conditions. (**B, C**) Error bars indicate 95% confidence intervals; between-group comparisons (NR vs DR) were done using Mann-Whitney U test with Bonferroni correction; and within-group comparisons between different trial types were done using Wilcoxon Signed-Rank test with Bonferroni correction. (* : $p \leq .05$; ** : $p \leq .01$; *** : $p \leq .001$, **** : $p \leq .0001$)

mice, indicating a weakened representation of visual stimuli. To our surprise, during auditory trials, auditory stimuli explained more variance than motor activity in V1 and dorsal stream areas (Extended Data Fig. 2E) of both dark-reared and normally reared mice. At the same time, motor activity of dark-reared animals explained less variance in sound-evoked activity in V1, medial dorsal stream and ventral stream neurons compared to normally reared animals. To disambiguate these effects, we explored the unique contribution of auditory stimuli and motor activity during audiovisual trials (Fig. 3B), which constituted most of the experimental (and also model training and testing) data and during which, our model performed best (Extended Data Fig. 2C). In V1 and dorsal stream neurons of dark-reared mice, we found auditory stimuli to explain more variance in audiovisual responses than in normally reared mice. Crucially, auditory stimuli explained more variance than motor activity in audiovisual responses, but only in V1 and dorsal stream neurons of dark-reared mice. For normally reared mice, the auditory and motor contribution to audiovisual responses was equal,

except in V1 where motor activity explained slightly more variance than auditory stimuli. V1 was also the only region where motor activity explained more audiovisual response variance in normally reared compared to dark-reared animals, while in the higher visual regions there was no difference between the rearing conditions. A follow-up analysis revealed distinct sets of neurons whose neural responses to audiovisual stimuli were best explained by visual stimulus features, auditory stimulus features, or motor activity (Fig. 3C). Across the visual cortex, dark-reared mice had a lower proportion of neurons whose activity was best explained by visual stimulus features than normally reared mice. V1 and dorsal stream of dark-reared mice contained a higher proportion of neurons whose audiovisual responses were best explained by auditory stimulus features compared to normally reared mice. Furthermore, the ventral stream of dark-reared mice contained higher proportion of neurons best explained by motor activity relative to normally reared mice. Together these results suggested that dark

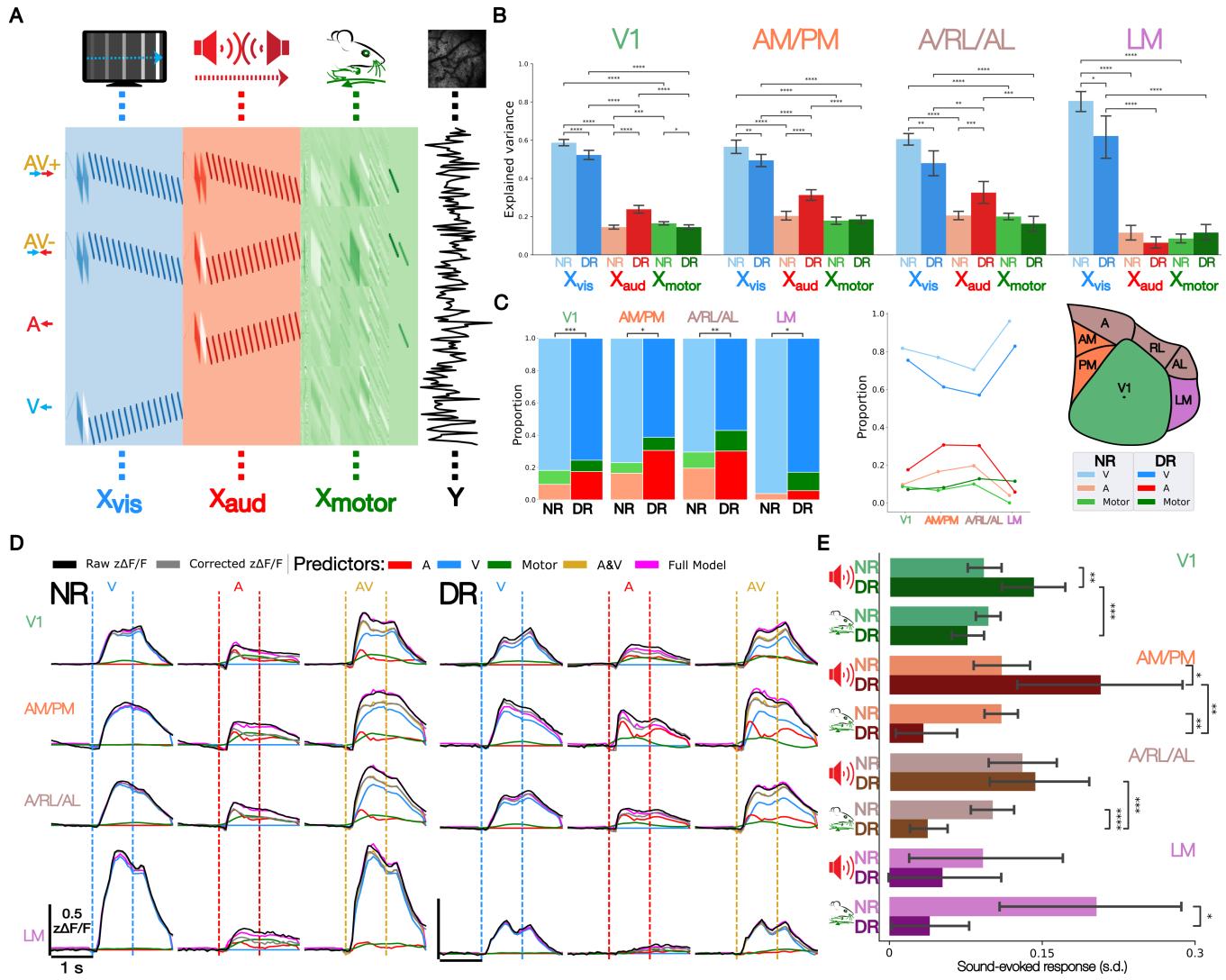


Figure 3: Dissociation of sensory and motor-related neural activity reveals weaker representation of vision and enriched representation of sounds in V1 and dorsal stream neurons of dark-reared mice. **A)** Regression model. Single-neuron firing rate (Y) was modeled as a linear combination of three sets of predictors (and a single trial-number control predictor; not shown) stored in a (*recording frames* \times *predictors*) design matrix (X). Visual predictors (X_{vis}) and auditory predictors (X_{aud}) contained information about presence, direction and location of the moving bar and moving sound respectively. Motor predictors (X_{motor}) contained raw, square, derivative and thresholded motion onset information about running speed, whisker motion energy and pupil size movement signals. Example shows four different consecutive trial types with various stimulus directions. **B, C)** Explained variance of audiovisual trial-averaged single-neuron activity across predictor sets, brain areas and rearing conditions. **C)** Proportion of neurons with maximum explained variance of audiovisual trial activity coming from a given predictor set across brain areas and rearing conditions. Fisher's exact test (NR vs DR). **D)** Average single-neuron responses by brain area across trial types and rearing conditions decomposed into additive contributions of predictor subsets. Corrected signal (grey) = Raw signal (black) - Predicted motor signal (green). Neurons with significant explained variance by the full model for held-out trials of any type were used (NR: 2266 neurons; DR: 1349 neurons). **E)** Predicted sound-evoked fluorescence response using auditory versus motor predictors across brain areas and rearing conditions. Neurons with significant explained variance by the full model for held-out auditory trials were used (NR: 747 neurons, DR: 378 neurons). **(B, C)** Explained variance was calculated on 20% held-out trials. Neurons with significant explained variance by the full model for held-out audiovisual trials were used (NR: 1801 neurons, DR: 948 neurons). **(B, E)** Error bars indicate 95% confidence intervals; between-group comparisons (NR vs DR) were done using Mann-Whitney U test with Bonferroni correction; and within-group comparisons between different predictor sets were done using Wilcoxon Signed-Rank test with Bonferroni correction. (* : $p \leq .05$; ** : $p \leq .01$; *** : $p \leq .001$; **** : $p \leq .0001$)

rearing weakens representation of visual information across the visual cortex and enhances representation of sounds in single-neuron responses of V1 and dorsal stream neurons.

To investigate the timecourse and response strength of the auditory-related and motor-related responses predicted by our model, for each brain area and rearing condition, we obtained the average single-neuron responses predicted by different sub-sets of predictors (Fig. 3D). In V1 and dorsal stream regions, the sound-evoked responses predicted by auditory components largely preceded the sound-evoked responses predicted by motor components across rearing conditions (Fig. 3D, Extended Data Fig. 2F), although this trend was somewhat less pronounced in dark-reared animals. When comparing the response strength of auditory and motor components of sound-evoked responses we found stronger auditory-related responses in dark-reared animals in V1 and the medial dorsal stream and weaker motor-related responses across the higher visual areas relative to normally reared animals (Fig. 3E). Furthermore, in V1 and across dorsal stream areas, auditory-related responses were stronger than motor-related responses, again only in the dark-reared group. In normally reared animals, auditory-related and motor-related sound-evoked responses did not differ in response strength across visuocortical regions. These results suggest that in V1 and dorsal stream regions of dark-reared mice, the early auditory-related response dominates over the later motor-related sound-evoked response; whereas the auditory-related and motor-related responses to sounds are equal in strength across the visual cortex of normally reared mice.

2.3 Dark rearing alters the distribution of unimodal and multimodal neurons across spatial scales in the visual cortex

After removing the motor-related components from all recorded single-neuron activity (see Fig. 3D for average corrected single-neuron responses), we proceeded to investigate single-neuron activity evoked by visual, auditory and audiovisual stimuli in dark-reared and normally reared mice. Focusing on stimulus-responsive neurons, we found functionally distinct, non-overlapping subpopulations of visual, auditory and multimodal neurons across the visual cortex of both dark-reared and normally reared mice (Fig. 4A-B, Extended Data Fig. 4). Investigating the distribution of these functional neuron types across visuocortical regions, we found that V1 and dorsal stream regions of dark-reared mice contained a lower proportion of visual neurons and higher proportion of multimodal neurons compared to normally reared mice (Fig. 4C). In both rearing conditions, the ventral stream contained the highest proportion of visual and lowest proportion of auditory and multimodal neurons. V1 and dorsal stream regions of normally reared mice had nearly identical neuron type composition ($\pm 50\%$ visual, 30% multimodal and 20% auditory). In dark-reared mice, V1 and medial dorsal stream contained a high proportion of multimodal neurons ($\pm 30\%$ visual, 50% multimodal and 20% auditory), while the proportion of visual and multimodal neurons was balanced in the lateral dorsal stream ($\pm 40\%$ visual, 40% multimodal and 20% auditory). This shows that dark rearing leads to a shift toward nonvisual neuron-type composition in V1 and dorsal stream regions.

Next, we examined whether the differences in neuron type composition between visuocortical regions and rearing conditions, reflected more general spatial gradients, such as greater incidence of auditory neurons closer to lateral [19, 25] or anterior [26] borders of the visual cortex. Through bootstrapping, for each rearing condition we obtained a null distribution of spatial statistics reflecting the group-specific neuron type proportions and sampling location bias. Comparing real spatial statistics to the null distribution, we were able to find spatial patterns in the localization of different neuron types that could not be explained by the overall neuron type proportions or sampling location bias (Fig. 4D-F). Auditory neurons concentrated in the central portion while multimodal neurons concentrated in both the anterior and posterior portions of the anterior-posterior axis in normally reared mice (Fig. 4D). In dark-reared mice, this trend was effectively reversed, with auditory neurons concentrated in the posterior region (appearing less than expected in the central region), and instead multimodal neurons concentrating in the central region. In both dark-reared and normally reared animals, the concentration of visual neurons increased when approaching the lateral-most segments and in normally reared animals this was met with a decrease in proportion of auditory neurons (Fig. 4E). Additionally, in dark-reared mice, multimodal neurons significantly concentrated around medial and central bins along the mediolateral axis. As a composite measure of proximity to non-visual regions in all directions, we also looked at the distance from center of V1 and expected to see higher concentrations of auditory and multimodal neurons at greater V1 center distance (indicating proximity to nonvisual regions). To our surprise, normally reared mice showed a higher-than expected concentration of multimodal neurons around the center of V1, and higher than expected concentration of visual neurons (accompanied by lower than expected proportion of auditory neurons) far from V1 center (Fig. 4F). In contrast, dark-reared mice showed significant concentrations of auditory neurons in the center of V1, as well as, far from V1 center. At the same time, dark-reared mice showed lower than expected proportion of visual neurons in the center and far from the center of V1. Taken together, these results indicate that apart from region-specific differences in neuron type composition, dark rearing also changes the distribution of unimodal and multimodal neurons along multiple spatial directions in the mouse visual cortex.

Lastly, we wanted to see whether the three functional types of neurons we identified formed modality-specific like-to-like local clusters [17] across different brain regions and whether this was affected by visual experience. Drawing inspiration from previous work that used nearest neighbor analysis to analyze similar microclusters in area RL [17], to reduce the inherent noise in the single-nearest neighbor metric, we devised a "local neighborhood index". The cluster ("local neighborhood") identity of each neuron was determined by weighing the identity of neighboring neurons by their distance from each neuron in consideration (Fig. 4G). For each neuron type, we then determined the probability that it lies in a local neighborhood of a certain type across brain regions and rearing conditions (Fig. 4H). In V1 of normally reared mice, we found all three neuron types to participate in auditory neighborhoods more than expected by chance, whereas in dark-reared animals, auditory V1 neurons clustered with other auditory neurons and multimodal neurons clustered with multimodal neurons. In the medial dorsal

stream region, auditory and multimodal neurons of normally reared mice appeared in auditory neighborhoods above chance level, with multimodal neurons also clustering with other multimodal neurons. In dark-reared mice, multimodal neurons in the medial dorsal stream appeared in more auditory neighborhoods than expected by chance. The lateral dorsal stream region showed clustering of

visual and auditory neurons in visual neighborhoods across rearing conditions. While multimodal neurons of normally reared animals clustered together with other multimodal neurons, in dark-reared mice multimodal neurons appeared in visual neighborhoods above chance level. Finally, in area LM of normally reared mice, visual and multimodal neurons clustered in visual neighborhoods while

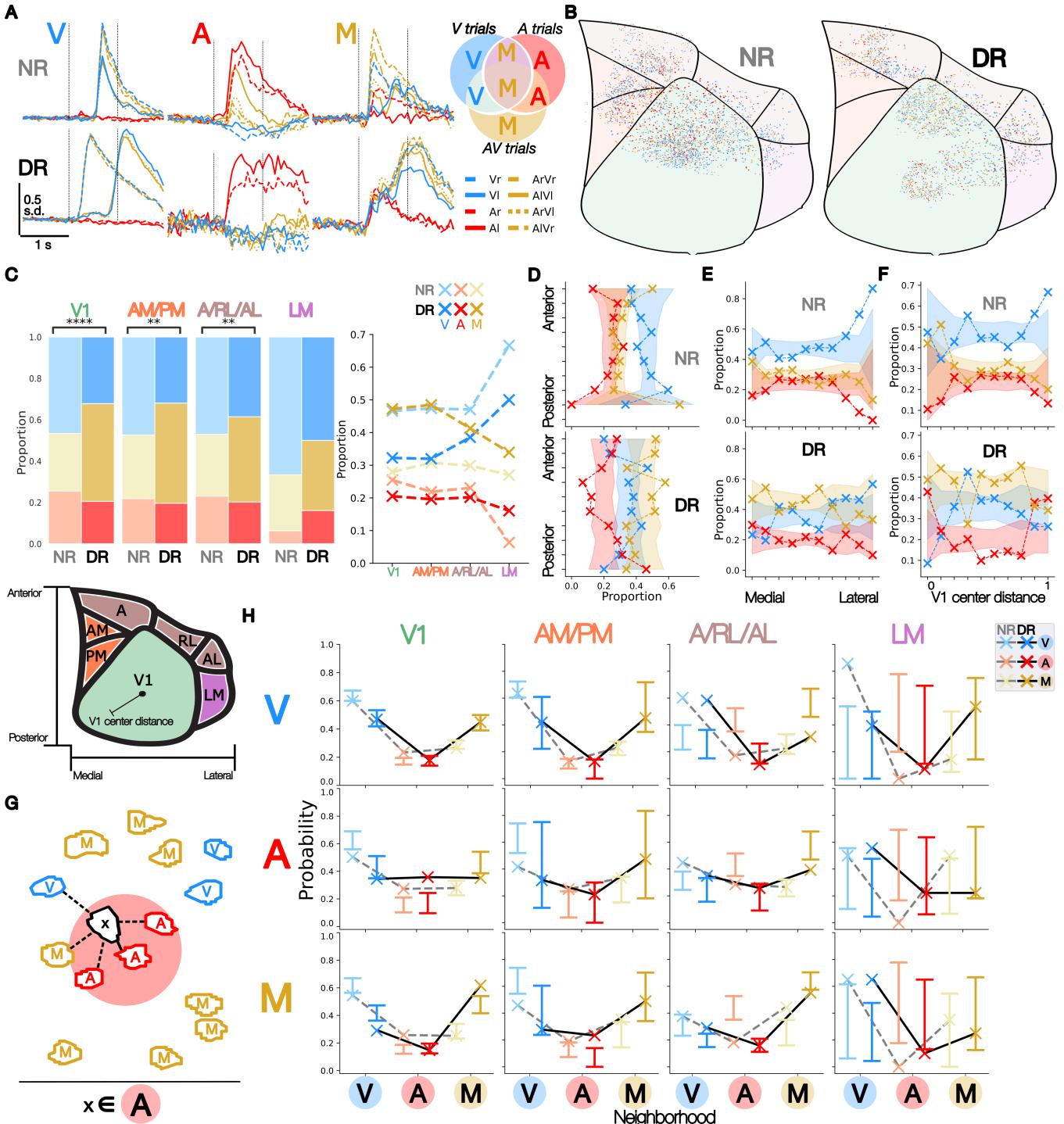


Figure 4: Dark rearing alters the distribution of unimodal and multimodal neurons across spatial scales in the visual cortex. **A)** Session-averaged single-neuron responses of example responsive neurons of each neuron type (Unimodal visual - V, Unimodal auditory - A, Multimodal - M) for both rearing conditions across trial types. **Top right:** Illustration of trial-type responsiveness inclusion criteria for each neuron type. **B)** Distribution of neuron types across visual cortex by rearing condition (NR: 1051 V, 528 A, 670 M; DR: 476 V, 275 A, 616 M). Nonresponsive neurons in grey (NR: 619, DR: 897). **C)** Proportion of neurons types across brain areas and rearing conditions. Fisher's exact test (NR vs DR) (** : $p \leq .01$; *** : $p \leq .0001$). **D)** Proportion of neurons along segments of anterior-posterior axis by neuron type across rearing conditions. **E)** Proportion of neurons along segments of mediolateral axis by neuron type across rearing conditions. **F)** Proportion of neurons at different distances from center of V1 by neuron type across rearing conditions. **G)** Local neighborhood index computation. For each neuron (neuron x, black), distances to k-nearest neighbors ($k = 5$) from the same recording session were computed (dashed and solid lines, solid line = nearest neighbor). The sum of neighbors of each neuron type in the k-nearest neighbors, weighed by the inverse of their distance to the neuron x was used to calculate "neighborhood weights" for neuron x and each possible type of neighborhood. The neighborhood type with the highest neighborhood weight was identified as the local neighborhood (auditory neighborhood for the example neuron x). **H)** Probability of being in a local neighborhood ($k = 5$) of each type (x-axis) across neuron types (rows), brain areas (columns) and rearing conditions. **(D-H)** Shaded regions or error-bars indicate the 95 % confidence interval of a 1000 times shuffled null-distribution (for the null hypothesis that spatial patterns in neuron position are determined by the overall proportion of each neuron type and/or the bias in sampling location.)

auditory neurons appeared in multimodal neighborhoods above chance level. In contrast, in dark-reared mice, auditory and multimodal neurons clustered in visual neighborhoods. Thus, on top of differences in neuron type composition, different regions across the visual cortex employ different microclustering schemes based on neuron type, some of which are affected by dark rearing.

2.4 Dark rearing affects stimulus responsiveness, direction selectivity, and multisensory integration in visuocortical neurons

The regression modeling and neuron-type composition analyses provided complementary evidence that dark rearing induces a shift from visual toward auditory processing at the level of individual neurons in V1 and dorsal stream regions. We wanted to know how this dark-rearing-induced shift affects the responsiveness and direction selectivity of the various functional subtypes of neurons we identified to stimuli of different modalities. At the same time, we wanted to know how these effects influence audiovisual integration on single-neuron level. We initially intended to also investigate the role of direction congruency in visual and auditory stimulus pairings, but finding no difference between the between congruent and incongruent audiovisual pairings across neuron types, brain areas and rearing conditions (Extended Data Fig. 5A-D), we concluded that the responses of the vast majority of neurons we recorded were not affected by our congruency manipulation. For that reason, in the following analyses, we did not split responses to audiovisual trials by congruency.

Unimodal visual neurons. Dark rearing reduced response amplitude of unimodal visual neurons to visual and audiovisual stimuli across V1 and HVAs (Fig. 5A-B). In V1, visual neurons on average responded more strongly to audiovisual trials than visual trials in both rearing conditions. In contrast, visual neurons in the dorsal stream of dark-reared mice, on average responded less to audiovisual trials than visual trials. Across visuocortical regions and rearing conditions, visual neurons were more direction selective for moving bars than moving sounds, confirming the "visual" identity of this group

of neurons (Fig. 5C). Additionally, visual neurons in V1 and medial dorsal stream of dark-reared mice were significantly less direction selective for visual stimuli compared to normally reared mice. Next, to assess audiovisual integration on single-neuron level, we quantified the response change and absolute magnitude of response change between response to the preferred visual direction and its maximally exciting sound pairing direction, as well as response change between non-preferred (opposite direction to preferred) visual direction and its maximally exciting sound pairing direction. In visual neurons across visuocortical regions of normally reared mice, pairing of nonpreferred bar directions with sound led to greater absolute response change than pairing the preferred bar direction with sound (Fig. 5D). In dark-reared mice, sound pairings with nonpreferred visual direction only elicited greater magnitude of response change in V1 and lateral dorsal stream visual neurons. A closer look at the signed response change in unimodal visual neurons revealed that pairing sound with preferred visual directions led to a balance of weak response enhancement and moderate response suppression, marked by low to negative response change index across visual areas and rearing conditions (Extended Data Fig. 6A). In contrast, pairing of sound with nonpreferred visual directions led to moderately strong response enhancement across visual areas and rearing conditions. Unexpectedly, visual V1 neurons of dark-reared mice displayed significantly greater response enhancement (and absolute response change) compared to visual V1 neurons of normally reared mice, when nonpreferred visual directions were paired with sound. Thus, dark rearing reduced the responsiveness of unimodal visual neurons to visual and audiovisual stimuli across the visual cortex, while only reducing visual direction selectivity in V1 and medial dorsal stream. Unimodal visual neurons showed greater response change in audiovisual trials with nonpreferred rather than preferred visual directions, and in V1 visual neurons dark rearing increased the multisensory enhancement during audiovisual trials with nonpreferred visual directions.

Unimodal auditory neurons. Unlike unimodal visual neurons, unimodal auditory neurons generally responded less to audiovisual stimuli compared to auditory stimuli (Fig. 5E-F). The sound-evoked responses of V1 auditory neurons of dark-reared mice were weaker

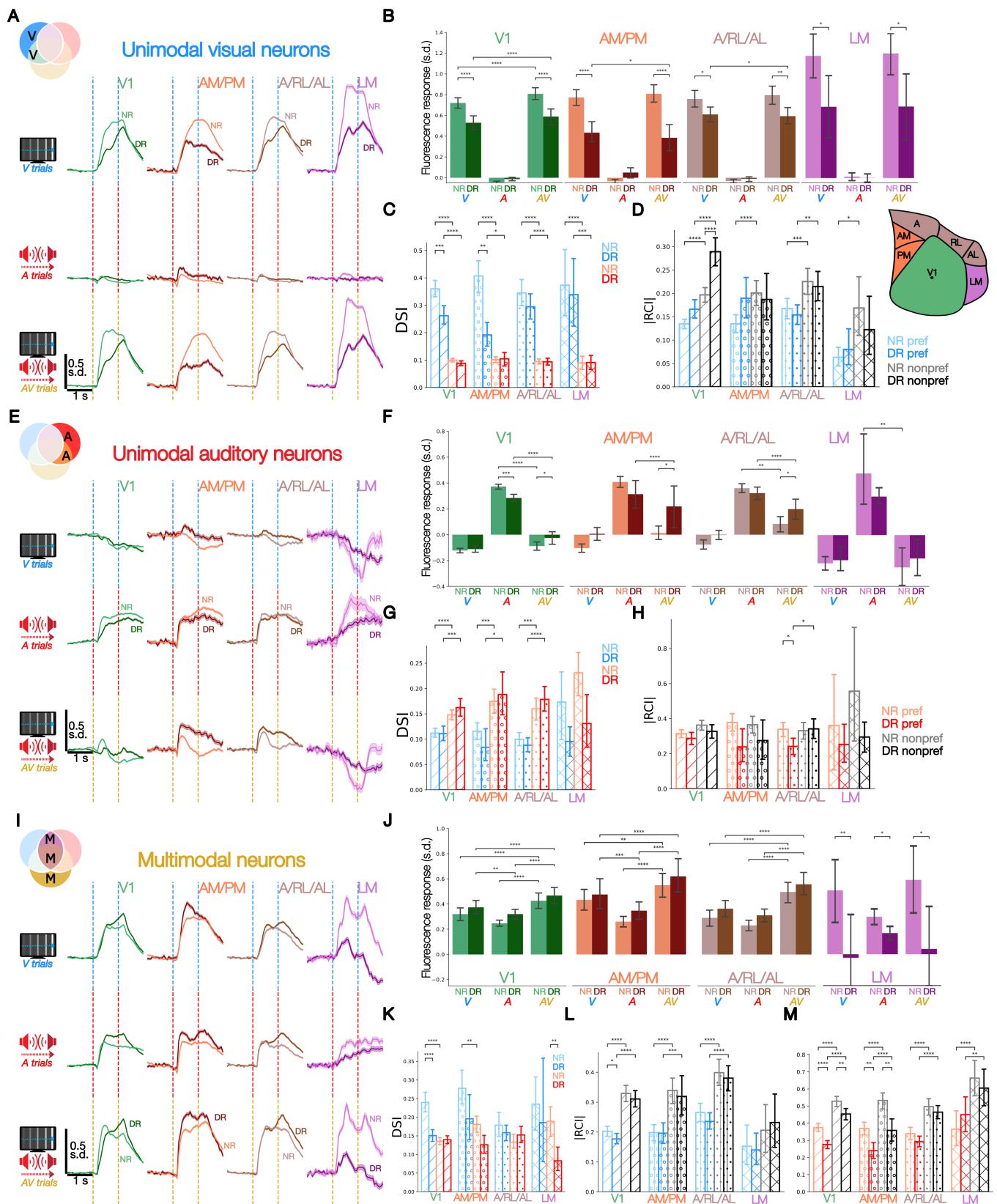


Figure 5: Dark rearing affects stimulus responsiveness, direction selectivity, and multisensory integration in visuocortical neurons. **A)** Average response of unimodal visual neurons across trial types (rows), brain areas (columns) and rearing conditions. For each neuron and trial type, the maximally stimulating stimulus direction was used. **B)** Stimulus-evoked fluorescence response of unimodal visual neurons across trial types, brain areas and rearing conditions. For each neuron and trial type, the maximally stimulating stimulus direction was used. **C)** *Direction selectivity index* (DSI) calculated for visual (blue) and auditory (red) stimuli of unimodal visual neurons across brain areas and rearing conditions. **D)** Absolute *Response change index* ($|RCI|$) calculated for pairing of a preferred visual direction with a sound (blue) and pairing of nonpreferred visual direction with a sound (grey/black), in unimodal visual neurons across brain areas and rearing conditions. Absolute value of RCI was used to measure magnitude of multisensory integration. **E)** Same as A) but for unimodal auditory neurons. **F)** Same as B) but for unimodal auditory neurons. **G)** Same as C) but for unimodal auditory neurons. **H)** $|RCI|$ calculated for pairing of a preferred sound direction with a moving bar (red) and pairing of nonpreferred sound direction with a moving bar (grey / black), in unimodal auditory neurons across brain areas and rearing conditions. **I)** Same as A) but for multimodal neurons. **J)** Same as B) but for multimodal neurons. **K)** Same as C) but for multimodal neurons. **L)** Same as D) but for multimodal neurons. **M)** Same as H) but for multimodal neurons. **(A, E, I)** Shaded regions indicate ± 1 SEM. **(Bar plots)** In all bar plots, error bars indicate 95% confidence intervals; between-group comparisons (NR vs DR) were done using Mann-Whitney U test with Bonferroni correction; and within-group comparisons between different trial types (B, F, J), modalities (C, G, K) or preferred and nonpreferred directions (D, H, L, M) were done using Wilcoxon Signed-Rank test with Bonferroni correction. (* : $p \leq .05$; ** : $p \leq .01$; *** : $p \leq .001$, **** : $p \leq .0001$)

compared to normally reared mice. In contrast, the responses of dorsal stream auditory neurons to audiovisual stimuli were enhanced by dark rearing. Across rearing conditions, V1 and dorsal stream unimodal auditory neurons were weakly, but significantly more direction selective for moving sounds than moving bars (Fig. 5G), confirming the "auditory" identity of these neurons. There was, however, no difference in sound direction selectivity of auditory neurons between rearing conditions. To study audiovisual integration in unimodal auditory neurons, we quantified the absolute response change magnitude when pairing preferred and nonpreferred sound directions with moving bars. In lateral dorsal stream auditory neurons of dark-reared mice, we found a lower response change magnitude to visual bar pairings with preferred sound direction relative to normally reared mice, as well as, relative to the response change of the same cells in audiovisual trials with nonpreferred sound directions (Fig. 5H). Inspection of the signed response change showed that pairings of a moving bar with the preferred sound direction, result in response suppression in both dark-reared and normally reared mice (Extended Data 6B). This response suppression was less pronounced in dorsal stream neurons of dark-reared mice. Similarly to visual neurons, pairings of visual stimuli with nonpreferred sound directions lead to more response enhancement and this effect is more pronounced in the medial dorsal stream of dark-reared mice compared to normally reared mice. In summary, dark rearing increased the responsiveness of dorsal stream unimodal auditory neurons to audiovisual stimuli, but had no effect on sound direction selectivity. Unimodal auditory neurons of dark-reared mice exhibited reduced multisensory suppression during audiovisual trials with preferred sound directions across the dorsal stream, but increased multisensory enhancement during audiovisual trials with nonpreferred sound directions in the medial dorsal stream.

Multimodal neurons. Across rearing conditions, multimodal neurons in the V1 and dorsal stream had stronger responses to audiovisual stimuli compared to either visual or auditory stimuli alone (Fig. 5I-J). Additionally, the responses of ventral stream multimodal neurons of dark-reared mice were weaker compared to normally reared mice. Visual and sound direction selectivity of multimodal

neurons generally did not differ, apart from V1 and medial dorsal stream regions of normally reared mice, which showed greater visual direction selectivity (Fig. 5K). Furthermore, in dark-reared mice, V1 multimodal neurons had lower visual direction selectivity and ventral stream multimodal neurons had lower sound direction selectivity compared to normally reared mice. When investigating audiovisual integration in multimodal neurons, we quantified the response change relative to each neuron's preferred visual and auditory direction separately. Multimodal neurons exhibited the same trend observed in visual and auditory neurons, namely greater response change in audiovisual pairings with the nonpreferred stimulus direction compared to the preferred stimulus direction (Fig. 5L-M, Extended Data Fig. 6C-D) across brain areas and rearing conditions. This was true both when pairing sounds with the (non)preferred visual direction (Fig. 5L, Extended Data Fig. 6C) and when pairing moving bars with the (non)preferred sound direction (Fig. 5M, Extended Data Fig. 6D). However, unlike in visual and auditory neurons, the dominant mode of audiovisual integration across regions and rearing conditions was response enhancement, for audiovisual trials with both preferred and nonpreferred stimulus directions (Extended Data Fig. 6C-D). Multimodal V1 neurons of dark-reared mice showed less pronounced response change in audiovisual trials with preferred visual directions (Fig. 5L) relative to normally reared mice. Similarly, V1 and medial dorsal stream multimodal neurons of dark-reared mice exhibited lower magnitude of response change during audiovisual pairings of both preferred and nonpreferred sound directions with moving bars (Fig. 5M) compared to normally reared mice. In conclusion, dark rearing reduced the responsiveness of ventral stream multimodal neurons to all stimulus modalities, and reduced visual direction selectivity in V1 multimodal neurons, as well as sound direction selectivity in ventral stream multimodal neurons. V1 multimodal neurons of dark-reared mice showed decreased response change in audiovisual trials with both preferred and nonpreferred sound directions, as well as, in audiovisual trials with preferred visual directions. Similarly, dark rearing resulted in reduced response change during audiovisual trials with both preferred and nonpreferred sound directions in medial dorsal stream multimodal neurons.

3 DISCUSSION

In this work, we explored the effect of early visual experience on audiovisual processing of individual neurons across the visual cortex. Partially aligned with our dorsal stream vulnerability hypothesis [15, 16, 26], we found that lack of early visual experience led to a shift away from visual processing across visuocortical regions and toward auditory processing in V1 and dorsal stream neurons. Likewise, single-neuron multisensory integration selectively depended on early visual experience in V1 and dorsal stream regions.

3.1 Lack of early visual experience biases V1 and dorsal stream neurons toward less visual and more auditory processing

The idea that sensory deprivation can lead to partial functional repurposing of the deprived modality sensory cortices by the intact modalities is supported by decades of research [1, 12, 26, 37, 38]. At the same time, the nature of sound-evoked activity in mouse visual cortex has been a subject of debate, with evidence for both low-dimensional movement-related origin [31], as well as, direct auditory cortex origin [32, 39]. Extending the findings of a previous study [32], we found that in addition to V1, the sound-evoked response of dorsal stream neurons can also be dissociated into early auditory-related and later movement-related components. In absence of early visual experience, the early auditory-related response to sound dominated over the later movement-related response in V1 and dorsal stream neurons. In contrast, neurons across the visual cortex of animals with normal visual experience showed balanced auditory and motor contributions to sound-evoked activity, indicating that neurons in V1 and dorsal stream regions might become more driven by auditory cortex input with lack of visual experience.

Based on a recent study [26] and the dorsal vulnerability hypothesis [15, 16], we expected lack of early visual experience to lead to a shift away from visual and toward auditory representations selectively in dorsal stream neurons. While we obtained dorsal stream evidence that auditory representations are enhanced and visual representations are weakened with lack of early visual experience, we also found the same effects in V1. This suggests that similar to dorsal stream, V1 neurons might also rely on experience-dependent mechanisms to establish an adaptive balance between feedforward visual and top-down auditory input. This idea was corroborated by our findings that in both V1 and dorsal stream regions, early visual experience establishes a balance between visual, auditory and multimodal neurons. Indeed V1 and dorsal stream regions of vision-deprived mice were marked by a less visual and more multimodal neuron-type composition. Additionally, we found that neurons in V1 and dorsal stream require visual experience to fully establish visual direction selectivity, providing further evidence for experience-dependent visual plasticity in V1 and dorsal stream alike.

Throughout this work we have referred to functionally distinct neuron "types", but crucially, each neuron's modality preference might itself be subject to experience-dependent plasticity. While unimodal neurons likely emerge through pruning and reinforcement of inputs from various modalities in clear favor of one modality; multimodal neurons may simply represent cases where feed-forward visual and top-down auditory cortex inputs are weighed

more equally. From this perspective, the increased representation of auditory information and higher proportion of "multimodal" neurons in V1 and dorsal stream might be viewed as a result of insufficient experience-dependent pruning of auditory cortex inputs on visuocortical "visual" neurons. While our study does not provide direct anatomical evidence of this mechanism, tracing studies of projections to visual cortex support this hypothesis [37, 40].

How can we account for our results consistently pointing to both V1 and dorsal stream, while previous research [26] showed experience-dependent audiovisual plasticity was largely restricted to the dorsal stream? An advantage of our experimental design is that we were able to compare each neuron's responses to intermixed visual, auditory and audiovisual stimuli within the same experimental session. This allowed us to divide neurons based on their preferred modality and thus isolate the effects of early visual experience on unimodal and multimodal neurons across brain regions. In contrast, previous research [26] studied visual and auditory plasticity in separate experiments. With no way to distinguish the neurons' preferred modality, the researchers inadvertently combined unimodal and a sub-set of multimodal neurons in one analysis in each of their experiments. Our data shows that lack of visual experience results in a more multimodal visual cortex composition at the cost of unimodal visual neurons; not necessarily accompanied by a higher incidence of unimodal auditory neurons. Thus, it is possible that some of the effects we were able to capture with our experimental design, perhaps including the V1 effects, were only detectable because we could distinguish between the modalities to which each neuron responded; and this was not previously possible with separate experiments for separate modalities. Another difference is that in the present study, we dissociated movement-related and sensory-related neural activity using a regression model, whereas Terra et al. [26] took a highly conservative approach and excluded neural data from all trials with stimulus-evoked movement above a fixed threshold. While excluding potentially problematic trials is a safe way to reduce false-positive rate, it can increase the false-negative rate in detecting stimulus-responsive neurons. Therefore, a second possibility is that by not dissociating movement-related activity from neural responses and simply excluding high-movement trials for all neurons (ignoring the relationship between each neuron's activity and the animal's movement), the previous study missed a significant portion of stimulus-responsive neurons in V1 (and potentially other areas as well).

3.2 Neuron-type-specific multisensory integration in V1 and dorsal stream regions is shaped by early visual experience

Having shown that early visual experience establishes a balance between visual and auditory processing, we asked how visual experience affects the integration of visual and auditory cues across the visual cortex. Based on a combination of human data from congenital cataract patients [9–11] and extensive evidence from superior colliculus of dark-reared cats [41–43], we expected lack of early visual experience to result in weaker multisensory integration. In addition, we hypothesized that dorsal stream HVAs would be particularly affected due to their developmental vulnerability and delay

[15, 16]. Our results showed that the nature of multisensory integration in the visual cortex is differentially shaped by visual experience based on individual neurons' feature preferences and sensitivity to inputs from visual and auditory modalities. Across the visual cortex, we observed the phenomenon of "*inverse effectiveness*", which has been characterized in the cat superior colliculus [4, 44]. Inverse effectiveness is defined as greater multisensory enhancement of stimuli that only weakly activate the neuron compared to stimuli that strongly activate the neuron when presented alone [4]. In the context of our experiment, this meant that multisensory enhancement was greater when pairing the less exciting (nonpreferred) stimulus direction with the other modality. Unimodal neurons generally showed response suppression when their preferred stimulus direction was paired with the other modality, but response enhancement when the nonpreferred stimulus direction was paired with the other modality. In V1 and dorsal stream neurons, lack of early visual experience increased the response enhancement observed in unimodal visual and unimodal auditory neurons during pairing of weakly stimulating bar and sound directions with the other modality. At the same time, absence of visual experience damped the response suppression of dorsal stream auditory neurons when preferred sound directions were paired with a moving bar. On the other hand, multimodal neurons were characterized by widespread response enhancement during audiovisual trials, regardless of stimulus preference (although still more pronounced for nonpreferred stimulus directions). This multisensory response enhancement was generally less pronounced in V1 and dorsal stream regions of mice reared without visual experience. In summary, these findings suggest that experience-dependent audiovisual plasticity in V1 and dorsal stream plays an important role in establishing single-neuron integration of visual and auditory cues.

How can we conceptualize the interplay between multisensory integration and functional neuron types? The distinct effects of visual experience on unimodal and multimodal neuron subpopulations suggest that in addition to establishing the right balance between feedforward visual and top-down auditory input, early experience of cross-modal environmental statistics might directly shape local circuits involved in multisensory integration. Our results show that unimodal neurons are suppressed while multimodal neurons are enhanced during audiovisual trials with individual neurons' preferred stimulus directions. One possibility, is that through experience unimodal neurons develop as nodes that convey specialized information about their modality-specific tuning, while multimodal neurons establish connections allowing them to integrate information about stimulus features across modalities and send their computations further up the visual system. Although our experiment cannot tease apart the exact circuitry involved, a combination of enhancing [22, 39] and suppressing [21, 45] inputs from auditory cortex impinging on local interneuron circuits could be responsible for the observed effects. For instance, multimodal neurons might connect to a local network of interneurons that inhibits surrounding unimodal neurons upon activation of multimodal neurons by a specific combination of cross-modal features. At the same time, the principle of inverse effectiveness is thought to serve the general purpose of enhancing detection of weakly effective stimuli [4], and for that reason might be advantageous in unimodal and multimodal neurons alike.

3.3 Limitations and future directions

Taken together, the present findings enhance our understanding of experience-dependent audiovisual plasticity across the visual cortex, but they also expose several limitations that define clear avenues for further investigation. Against our expectations, we found no evidence for congruency-dependent audiovisual integration. One possibility is that our crude "simulated sound motion" approach did not succeed in activating the relevant neural circuits for sound localization and motion perception. Notably, our approach completely ignored interaural time differences and other key mechanisms used by the auditory system to localize sounds [46]. We found a small subset of unimodal auditory neurons that showed greater direction selectivity for sounds than for moving bars, suggesting that some visuocortical neurons could distinguish the simulated sound "direction". However, most neurons may have been completely unaffected by the simulated sound motion and instead only generally reacted to the white noise. Follow-up research could resolve this by decoding sound location and direction from responses of the various neuron types. Crucially, recent research shows that audiovisual integration in V1 does not depend on spatial coherence of auditory and visual stimuli, because auditory cortex inputs to V1 are not retinotopically mapped [39]. Hence, it is possible that even if we had used a speaker on a physically moving arm, there would be no effect of sound motion congruency, simply because visual neurons might receive information about sounds in random locations outside their receptive field. Still, evidence from V1 also suggests that audiovisual integration depends on the congruency between visual motion frequency and sound frequency [19]. Thus, auditory inputs in V1 might assist in integration of more complex audiovisual feature interactions than spatial location (which is handled by superior colliculus, where alignment of auditory and visual spatial maps aids audiovisual stimulus localization [47, 48]). Testing this hypothesis will require simultaneously presenting a variety of visual and auditory stimulus features (e.g. natural movies with natural sounds), to determine which cross-modal interactions are processed in the visual cortex.

Despite the proposed parcellation of the dorsal stream into medial and lateral divisions [33], individual dorsal stream areas might differ in their role in nonvisual processing. As such, future studies should aim to focus on all HVAs separately. This would elucidate whether dorsal stream sensitivity to early visual experience for audiovisual processing is predominantly driven by individual HVAs, and whether functional properties of individual HVAs determine their role in audiovisual integration. Although we did not find evidence of a general microclustering rule in any of the surveyed regions, clustering of unimodal neurons in lateral dorsal stream area RL has been reported in the visuotactile context [17]. Future analysis of microarchitecture of individual HVAs might therefore reveal trends that are lost in grouped analysis of HVAs.

Finally, future studies should causally investigate the local circuitry of unimodal and multimodal neurons in the context of multisensory integration. Computational modeling and perturbational experiments involving interneurons, as well as, top-down and subcortical crossmodal inputs to the visual cortex, will be necessary to discover the plasticity mechanisms and connectivity rules underlying the ability of neural circuits to integrate multiple senses.

4 METHODS

4.1 Animals

All animal experiments received approval from the institutional animal care and use committee of the Royal Netherlands Academy of Arts and Sciences under CCD license AVD80100202215934. The study involved both male and female mice, with distinct cohorts subjected either to dark-reared (8 mice) or normally reared (11 mice) conditions. Animals were generated as F1 offspring from crossings of female CBA/JRj mice (Janvier Labs) with male Vipr2-Cre^{-/-} mice (Jackson Laboratories, strain 031332). All animals underwent surgery around 60 days after birth and their first experimental recording session around 90 days after birth. Prior to surgery, mice were housed in groups, and post-surgery, they were maintained individually or in groups with unrestricted access to food and water, adhering to either a reversed 12-hour day/night cycle (normal rearing) or complete darkness (dark rearing). Running wheels were provided within all home cages. Care was meticulously taken to ensure that dark-reared mice experienced absolute darkness until their initial recording session, with daily inspections and habituation sessions conducted under infrared illumination (850 nm) using custom infrared binoculars. Surgical procedures involved protective covers placed over the eyes to prevent exposure to ambient light. All recordings took place during the dark phase for the normally reared group.

4.2 Viral injection and cranial window surgery

For surgical procedures, anesthesia was induced and maintained using Isoflurane (5% induction, 1.5–2% maintenance in oxygen), with body temperature held constant at 37°C with a heating pad and eyes covered with light-protective caps. Analgesic administration included Metacam (5 mg/kg subcutaneously), Temgesic (0.1 mg/kg subcutaneously), Xylocaine topical gel (applied on periost), and postoperative Rymadil (0.06 mg/mL, in drinking water for three days). Anti-inflammatory treatment consisted of Dexamethasone (4–8 mg/kg subcutaneously). Following removal of skin and periosteum above the skull, eight small holes were drilled into the skull at strategic locations across the visual cortex using a dental drill. Viral vectors comprised of 1:1 mixture of AAV1-hDlx-dlox-mCyRFP1(rev)-dlox (0.67 × 10¹² titer, VVF Zurich, v313-1) and AAV9-CaMKIIα-jGCaMP8m (0.67 × 10¹² titer, VVF Zurich, v630-9). Eight injections of 9.2 nL of the viral vector were injected at depths of 200 μm and 400 μm below the cortical surface, at each of the eight injection sites at a rate of 46 nL/s. Following a craniotomy above the visual cortex, a cranial window, constructed from layered coverslips (4+4+5 mm), was placed into the craniotomy, secured with dental cement along with a metal head fixation ring. Postoperative recovery lasted at least one week, followed by gradual habituation to head fixation in the recording setup, until comfortable running was observed.

4.3 Population receptive field mapping

To perform population receptive field mapping, the complete cranial window was imaged using a wide-field fluorescence microscope (Axio Zoom.V16Zeiss/Caenotec-Prof. Ralf Schnabel). A high-speed sCMOS camera (pro.edge 5.5) was used to capture the images at 20

Hz, with 50 ms exposure time (1600 x 1600 pixel resolution) and recorded with Encephalos software (Caenotec-Prof. Ralf Schnabel) after which the images were down-sampled to 800 x 800 pixel resolution for further population receptive field analysis.

During the wide-field receptive-field mapping session, mice were positioned at the center, 14 cm away, from a LCD screen covering a 144° x 86° field-of-view (122 x 68 cm iiyama LE5564S-B1 screen, 1920 x 1280 pixels at 60 Hz refresh rate). Stimuli were created and presented using COGENT (developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience) together with Matlab (Mathworks). A checkerboard pattern of eccentricity-corrected angled bars at 0°, 45°, 90° or 135° with 10° diameter was shown on a grey background (20 cd/m²). 58 stimuli were each pseudo-randomly presented 10 times for 0.5 s with a 3.6 s grey screen intertrial-interval (20 cd/m²). Wide-field images were smoothed using a Gaussian filter (SD = 2 pixels) and stored per trial. Population receptive field mapping and field sign analysis followed previously established protocols [49, 50].

4.4 Two-photon calcium imaging experiment

For two-photon calcium imaging, we used a Neurolabware 2P microscope with a Ti-sapphire laser (Mai-Tai ‘Deepsee’, Spectra-physics, running at 920 nm) together with a Nikon 16x 0.8 NA water-immersion objective (at 1.0x to 1.6x digital zoom). Imaging frames were captured at approximately 15.5 Hz, with the Scanbox software (Neurolabware) coordinating image acquisition. Mice were head-fixed and free to run on a wheel, with running speeds tracked using an Arduino-driven rotary encoder.

4.4.1 Audiovisual experiment The audiovisual experiment used moving bars and “simulated” moving sounds. Therefore, in any given trial a stimulus of either modality could: 1) move leftwards, 2) move rightwards, or 3) be absent. Because trials with both stimuli absent were not presented, there were 8 trial types in total. Trial types were presented in pseudo-random order, with each trial type being presented 90 times for a total of 720 trials in each two-photon audiovisual recording session.

Visual stimuli. During the audiovisual experiment, mice were centered 15 cm away from a 24-inch gamma-corrected HD LED monitor (1920 x 1080 pixels at 60 Hz refresh rate) covering a 120° x 90° field of view. Visual stimuli (generated and shown using OpenGI and Psychophysics Toolbox 3 in Matlab) consisted of a single high-contrast vertical moving bar (full screen height; 0.05 cycles per degree; 10° bar half-width; 55 % contrast; 6 cycles / s) that drifted horizontally across the full width of the screen in one of two possible directions in a given trial (rightward or leftward). Each presentation lasted 1 s, and was followed by a 2–3 s mean-luminance grey screen randomized duration inter-trial interval.

Auditory stimuli. During the same audiovisual experiment, while being centered in front of the screen, mice were simultaneously centered between two ultrasound speakers (Batsound, L400 Ultrasound speaker) positioned 20 cm away at ear level lateral to each ear. Auditory stimuli (generated and presented using Psychophysics Toolbox 3 in Matlab) consisted of gaussian white noise (band-pass filtered between 3–64 kHz) presented from each speaker simultaneously for 1 s (including a 5 ms cosine-squared rise/decay time),

followed by a 2-3 s randomized duration inter-trial interval with no auditory stimuli. To simulate sound motion via changing interaural intensity difference between the right and left speakers over the course of the trial (leftward or rightward motion), an exponential intensity ramp-up was applied to the speaker towards which sound was "moving" (e^{5t} with $0 \text{ s} \leq t \leq 1 \text{ s}$) and an exponential intensity ramp-down was applied to the speaker from which the sound "originated" in a given trial (e^{-5t} with $0 \text{ s} \leq t \leq 1 \text{ s}$). The intensity ranged between 0 - 70 dB SPL RMS. The auditory stimuli were presented at 192 kHz using the Terratec Aurean Xfire 8.0 HD external soundcard. Peak intensity of background noise was 40 dB SPL. Sound intensity was calibrated with a condenser ultrasound microphone (Avisoft-Bioacoustics CM16/CMPA) placed at ear level, together with UltraSoundGate 116Hb recording interface and Avisoft-RECODER USGH software.

4.4.2 Two-photon calcium imaging data preprocessing Raw two-photon images were pre-processed using the SpecSeg toolbox [51] to detect regions of interest (ROIs) corresponding to neuronal somatas and extract the mean fluorescence signal across the pixels of each ROI. Rigid motion correction was performed using NoRM-Corre [52], followed by automated ROI detection and selection based on cross-spectral power across pixels. Afterwards, ROIs were manually refined, and mean signal was extracted per ROI. Neuropil correction was applied by subtracting 70% of the average pixel values from a "doughnut shape" surrounding each ROI. $\Delta F/F$ values were calculated by subtracting a moving baseline and dividing by a linear fit of that moving baseline (10th percentile over 500 frames). $z\Delta F/F$ values were calculated by z-scoring the $\Delta F/F$ signal:

$$z\Delta F/F = \frac{\Delta F/F - \mu(\Delta F/F)}{\sigma(\Delta F/F)} \quad (1)$$

The standardized $z\Delta F/F$ signal was used in most analyses to standardize and align the units between neural and movement signals. All neural signals were trial-locked to extract -1 to +2 s around stimulus onset (1 s baseline, 1 s stimulus presentation, 1 s post-stimulus period), which yielded 47 samples per trial per neuron. For most analyses the $z\Delta F/F$ were also baseline-corrected on trial-level.

4.4.3 Matching two-photon ROIs to visual brain areas To identify the visual cortical areas in which each two-photon ROI resides, wide-field population receptive field maps were manually aligned to the Allen Brain Atlas [53]. Images of the two-photon field of view from each recording were manually aligned to the wide-field macroscope image of the whole cranial window for each respective mouse. Finally, using both alignments, individual two-photon imaging ROIs were mapped onto the Allen Brain Atlas coordinates and assigned visual brain area labels using Matlab.

4.4.4 Movement tracking To record pupil and whisker movements during two-photon calcium imaging, we used an infrared camera (FLIR, Flea3 USB3 combined with a Tamron 12VM412ASIR 1/2" 4-12mm F/1.2 Infrared Manual C-Mount Lens), triggered by two-photon image acquisition sampled at around 15.5 Hz. Using Facemap [54], we extracted the average motion energy signal from

a manually drawn region of interest around the whisker pad. Similarly, pupil size signal (corrected for corneal reflection) was extracted from a hand-drawn region of interest using Facemap. Using Matlab, the extracted movement signals (including the running speed signal extracted from a rotary encoder connected to the running wheel) were all trial-locked to extract -1 to + 2 s around stimulus onset, to yield 47 samples per trial.

To calculate standardized stimulus-evoked movement signals, we z-scored the trial-locked movement signals on the baseline (-1 to 0 s) mean and standard deviation calculated over the entire recording session, followed by baseline-correction on trial-level. This approach was chosen because attempts to z-score on trial-level frequently led to extreme inflation of movement signal amplitude in trials with low baseline variability, to the point that any further effects and analyses were primarily driven by these inflated trials.

4.5 Data Analysis

All analyses described below were performed using custom written Python 3.13.2 code, publicly available on GitHub (see [Analysis Code](#)).

4.5.1 Two-photon calcium imaging analysis

Fluorescence response calculation. To calculate the fluorescence response (FR) of each neuron in a given trial, we devised a custom calculation to meet the specific needs of our experiment. Initial attempts with a simple mean $z\Delta F/F$ over the 1 s stimulus presentation period in each trial showed that this naive method did not reliably capture neuronal responses to our stimuli. The reason is that because we used moving stimuli, and recorded from thousands of neurons without knowing their single-neuron receptive field locations, each neuron could start responding when the bar passed any point on the screen, including the final centimeters of the screen covered in the final stimulus presentation imaging frames in each trial. Combined with the relatively low temporal resolution of our 2P imaging setup (15.5 Hz) and the relatively slow kinetics of GCaMP8m, many neurons with clear responses starting later in the trial were mischaracterized as unresponsive when averaging over stimulus presentation frames.

To improve the fidelity of the FR calculation to the experimental data, we modified the calculation to capture mean of a 1 s window around peak of each neuron's response in each trial. Namely, for each neuron, in each trial we calculated the FR for three 1s windows (0-1 s, 0.5 - 1.5 s and 1.0 - 2.0 s). Then, for each neuron we tried to estimate whether it shows increased (median FR > 0) or decreased (median FR < 0) fluorescence responses across trials to the stimuli presented. Based on that we chose the maximum of the three calculated FRs in each trial across trials for neurons with median FR > 0, and chose the minimum of the three calculated FRs in each trial for neurons with median FR < 0. A neuron's FR for a given trial type was calculated as the average of its FRs across all trials of that trial type.

Detection of significantly responsive neurons. To identify significantly responsive neurons whose FRs to the stimulus (calculation described above) were significantly higher than FRs during the 1 s pre-stimulus baseline, we applied a permutation test. Specifically, a neuron was deemed significantly responsive if its FR for at least

one of the trial-types was higher than 99th percentile (or lower than 1st percentile for neurons with significantly suppressed firing) of a null distribution of baseline neuronal responses. The null distribution was obtained by calculating each neuron's FRs for the 1 s pre-stimulus window 1000 times with the $z\Delta F/F$ values circularly shifted by a random number of frames every shuffle.

Classification of unimodal and multimodal neurons. As shown in Fig. 4A (top right illustration), neurons were classified as unimodal (visual or auditory) if they were significantly responsive during at least one of the two directions of stimulus presentation for one sensory modality and not significantly responsive for any stimulus direction of the other sensory modality. Neurons were classified as multimodal if they were significantly responsive during at least one of the trial types with paired presentation of both modalities (audiovisual trials) and not significantly responsive during presentation of either modality alone; if they were significantly responsive during presentation of both modalities alone but not during audiovisual trials; or if they were significantly responsive for at least one direction in each trial type.

4.5.2 Movement signal analysis To quantify the magnitude of the movement signals in Fig. 2B and account for the different onset latencies, we applied the same procedure as described above for the fluorescence response calculation to extract a motor response value for each trial and each z-scored movement signal. For analysis in Fig. 2C, we calculated the population neural activity in each session as the average z-scored signal over all the ROIs recorded in a given session, yielding a (trials x timestamps) population activity matrix for each session. For each trial, we calculated the Pearson correlation coefficient during stimulus presentation frames (0-1 s) between each movement signal and population activity. For each session, we then obtained the average correlation during a specific trial type using the Fisher-z transformation.

4.5.3 Regression model To dissociate visual, auditory and motor components of neural activity, we built a linear regression model for encoding single-neuron activity [22, 32, 36] using stimulus-related and motor predictors.

Design matrix. As illustrated in Extended Data Fig. 2A, the horizontally moving stimuli were modeled using 16 location predictors, sectioning the horizontal space into a resolution of one location predictor per frame during which stimulus was presented. Bars simply moved across the screen with a constant contrast and orientation and therefore bar motion over the trial was modeled with a binary flag indicating whether the bar was at a certain location at a certain frame during the trial. A simplifying assumption was used that the bar only covered one location predictor at a time, while in reality, the entire bar covered 1/6th of the monitor screen. As such, our model can more accurately be seen as modeling the screen location of the center of the bar. On the other hand, sound motion was simulated by a simultaneous exponential ramp-up and ramp-down of sound intensity of two speakers positioned next to the animal's ears. As shown in Extended Data Fig. 2A, the result of this approach was that the overall sound intensity was highest at the beginning and end of each trial and lowest in the middle each trial. Thus, while maintaining the assumption that the sound only occupied one sound location predictor in a given recording

frame, the value of the location predictor reflected the overall sound intensity at that point in the trial. For each stimulus modality, an additional binary predictor was added that indicated the stimulus presence or absence across recording frames; as well as, another predictor that indicated stimulus direction. This resulted in 18 initial stimulus-related predictor features for each stimulus modality (16 location predictors, 1 presence predictor and 1 direction predictor). To account for physiologically plausible delays of neural responses to stimuli, as well as, the relatively slow nature of the fluorescence signal without inflating the parameter count, we used temporal basis functions, as shown in Extended Data Fig. 2B. Namely, we used nine raised cosine basis functions (0.3 s half-width) that spanned the lags between 0 and 1 seconds and that we convolved with all stimulus predictors to yield the final 324 stimulus predictors ($18 \times 9 = 162$ visual predictors and 162 auditory predictors). Motor predictors were constructed from the three standardized stimulus-evoked movement signals that we extracted as described above, namely running speed, whisker motion energy, and pupil size. For each movement signal, we obtained four initial predictor features: 1) raw movement, 2) squared movement, 3) derivative of movement, and 4) thresholded movement onset. Squared movement emphasized periods of high stimulus-evoked movement, while derivative of the movement signal captured the instantaneous change in movement evoked by stimulus. Onset terms were binary and in each trial where a movement "event" occurred, indicated the first moment after stimulus onset when the standardized stimulus-evoked movement exceeded 2 SD from baseline. Similarly to stimulus-related predictors, to account for temporal offsets between movement and neural activity, we convolved all motor predictors with 10 raised cosine basis functions (0.2 s half-width) that spanned both advanced and delayed temporal relationships (-0.3 to 0.5 seconds; negative lags indicating neural activity preceding motor predictors). This yielded a total of 120 motor predictors. To ensure no data leakage across trials, the basis functions and both stimulus and motor predictors were convolved on trial level. As a control, we also included a single trial-number predictor to absorb overall drift in fluorescence / activity over the course of the session.

Thus, for each experimental session, we constructed a tall, $T \times P$ design matrix X , where P represented the number of predictors (162 visual + 162 auditory + 120 motor + 1 trial number = 445 predictors). Because the model was fit on concatenated trial-locked data, T represented the number of trials in a session (720) multiplied by the number of recording frames in each trial (± 65 ms bins, -1s to +2s, 47 frames per trial).

Model fitting. Each single-neuron encoding model was fit with ridge regression, using the RidgeCV function from the Scikit-learn Python library [55]. For each single-neuron model, the optimal regularization parameter λ was estimated using leave-one-out cross validation. To prevent overfitting, the range of λ for cross-validation was restricted to medium-large values (0.6 to 10000). We chose ridge regression instead of other regularization methods, because we sought to split the contribution to neural activity between correlated sensory and motor predictor sets, instead of isolating the few best predictors and discarding other correlated variables.

Model evaluation. To assess model performance, we held-out data from 20% of the trials using a custom train-test split. Because

our experiment contained 90 repetitions of each possible stimulus configuration, with all trial types randomly intermixed in each session, holding-out every fifth trial of each trial type guaranteed a representative sample of trials obtained throughout the session for model testing. We quantified model performance using the explained variance (EV) metric computed on held-out trials:

$$EV = 1 - \frac{\text{var}(Y - \hat{Y})}{\text{var}(Y)} \quad (2)$$

where Y represents the single-neuron activity in held-out trials and \hat{Y} is the model-predicted activity for held-out trials. Each neuron's activity was deemed significantly explained by the model for a given trial type, if the model's held-out EV for that trial type exceeded the 99th percentile of a shuffled null distribution. After fitting to the train set, each neuron's null distribution was obtained by fixing the fitted weights and circularly shuffling the held-out neural signal 1000 times and recomputing the EV metric. This procedure destroyed the trial-type and temporal correspondence between held-out neuronal activity and the predictor design matrix X and resulted in a null distribution of the EV metric for a particular neuron.

We quantified the contribution of different predictor sets to neural activity with held-out predictor-set EV (EV_P):

$$\hat{Y}_P = X_P \beta_P, \quad P \in (\text{vis, aud, motor}) \quad (3)$$

$$EV_P = 1 - \frac{\text{var}(Y - \hat{Y}_P)}{\text{var}(Y)} \quad (4)$$

where X_P is a matrix containing the full predictor set P from the full design matrix X ; β_P is a vector of fitted coefficients for each predictor from the predictor set P ; and \hat{Y}_P is the activity predicted by the predictor set P . EV analyses were restricted to neurons with significant full-model EV for held-out trials of the relevant trial type.

Removal of motor components from neural activity. To remove motor components from neural activity, we wanted to achieve the best model fit possible for each neuron, and thus, we performed the model fitting procedure described above (leave-one-out cross-validated ridge regression) using both train and test sets combined. After fitting the model to the full session dataset for each neuron, we obtained the full-session trial-locked activity predicted by motor predictors (\hat{Y}_{motor}) as described above. Then, the corrected zΔF/F signal ($Y_{\text{corrected}}$) used in all analyses from Fig. 4 onwards, was obtained by subtracting the predicted motor-related activity from raw neural activity:

$$Y_{\text{corrected}} = Y - \hat{Y}_{\text{motor}} \quad (5)$$

4.5.4 Spatial distribution of neuron types After matching all ROIs from both rearing conditions to common coordinates within the Allen Brain Atlas map of mouse visual cortex [53] as described above, for each neuron we computed the following spatial statistics (min-max normalized across all neurons across rearing conditions and recording sessions):

- Position along the anterior-posterior axis, with 0 indicating most posterior and 1 indicating most anterior neuron location.

- Position along the medio-lateral axis, with 0 indicating most medial and 1 indicating most lateral neuron location
- Distance from center of V1, with 0 indicating most proximal and 1 indicating most distant neuron location to center of V1. Center of V1 was calculated as center of mass of the V1 region from the Allen Brain Atlas map of mouse visual cortex [53].

We split discretized each of the above spatial axes into 10 bins and for each group separately, computed the proportion of significantly responsive neurons in each bin corresponding to one of the three functional neuron types.

In Fig. 4D-F, we used a permutation test to determine which spatial patterns could not be explained by a bias in recording locations, overall neuron type proportions and chance alone. By randomly shuffling the neuron type labels between all significantly responsive neurons in each condition and recomputing the neuron type proportions for all bins 1000 times, we obtained a null distribution of neuron type proportions along each spatial axis, which reflected the overall neuron type proportions in each group, as well as, the rearing-condition-specific bias in recording locations. Bins with real-data neuron type proportions falling outside the 95% confidence interval of this null distribution for each neuron type and rearing condition, were considered to represent significant trends in low or high localization of a particular neuron type to a given spatial axis segment.

Local neighborhood analysis. To determine whether different neuron types form microclusters with the same or different neuron types, we devised a *Local neighborhood index*, illustrated in Fig. 4G. After embedding all neurons in a common coordinate system, as described above, for each session in each rearing condition, we obtained a matrix of pairwise distances between all neurons recorded in a given recording session. Because we were interested in clustering of the functional neuron types we found, we ignored nonresponsive cells in this analysis. We determined each neuron's local neighborhood (V / A / M neighborhoods) by a distance-weighted count of each neuron's k-nearest neighbors, attributing higher weight to closer neighbors than to distant neighbors ($w_{\text{neighbor}} = 1/d_{\text{neighbor}}$). Through experimentation using the elbow method on the average pairwise distance within a local neighborhood as k increases, we found $k = 5$ to be a good local neighborhood size (not shown). Each cell was thus assigned a local neighborhood based on neuron types of surrounding neurons from the same session. Using all responsive cells from a given rearing condition, for each neuron type, we then calculated the probability that cells of that neuron type will be part of each possible local neighborhood.

To determine whether certain neuron types occur in certain neighborhoods above chance, in Fig. 4H we again used a permutation test to construct a null distribution of probabilities for each neuron type to occur in any neighborhood type for each rearing condition. We obtained this null distribution by randomly shuffling the neuron type labels between all significant neurons from the same area within each recording session, and recomputing the local neighborhoods for all the neurons across rearing conditions 1000 times. Cases where certain neuron types occurred in one of the neighborhood types with a probability outside the 95% confidence

interval of the generated null distribution (different null distribution for each brain area \times neuron type \times neighborhood type \times rearing condition combination), were considered cases where neuron types clustered significantly above or below chance in a certain neighborhood.

4.5.5 Direction selectivity index We measured the strength of a neuron's direction tuning for both visual and auditory modality by calculating a direction selectivity index (DSI), which was adapted from a previously described orientation selectivity index [19, 56]. The advantage of this measure is that it explicitly accounts for trial-to-trial variability in neuronal activity. For each neuron and each modality M , we computed DSI as:

$$DSI_M = \frac{|\mu_{prefM}| - |\mu_{nonprefM}|}{\sqrt{(\sigma_{prefM}^2 + \sigma_{nonprefM}^2)/2}}, \quad M \in (V, A) \quad (6)$$

where μ_{prefM} and $\mu_{nonprefM}$ represent trial-averaged fluorescence response to the maximally, and minimally exciting stimulus direction during presentation of each modality M separately. Because for each modality our experiment only used two directions, the *non-preferred* stimulus direction was just the opposite of the *preferred* stimulus direction. The difference between, the average fluorescence response to the preferred vs nonpreferred stimulus direction was standardized by pooled variance of the fluorescence response distributions to the two stimulus directions. Because we were interested in the strength of a neuron's direction tuning, we used the absolute mean fluorescence responses, which resulted in a non-negative DSI index.

4.5.6 Response change index To study the magnitude and nature of multisensory integration, for each neuron we computed a response change index (RCI)[19] between its fluorescence response to isolated and paired presentation of visual and auditory modalities (M):

$$RCI_{M_p} = \frac{FR_{AVp} - FR_{Mp}}{FR_{AVp} + FR_{Mp}}, \quad M \in (V, A), p \in (\text{pref, nonpref}) \quad (7)$$

The index ranged between -1 and 1 with negative values indicating response suppression and positive values indicating response enhancement during audiovisual trials relative to visual (V) or auditory (A) trials alone. For each modality M , the metric was computed separately for audiovisual trials that used the preferred stimulus direction, as well as, for audiovisual trials that used the nonpreferred stimulus direction, thereby allowing us to study multisensory integration of weakly effective stimuli. For analysis of magnitude of multisensory integration, for each modality M and stimulus preference p , we analyzed the absolute RCI ($|RCI|$), which simply quantified how much audiovisual responses differed from responses to single-modality.

4.5.7 Statistical analysis Statistical comparisons are described in figure captions. Bootstrapping / permutation tests are described in detail in relevant sections above. All other statistical comparisons were carried out with two-sided, nonparametric tests with Bonferroni multiple comparison correction. Signal traces are always plotted as mean or mean \pm 1 SEM.

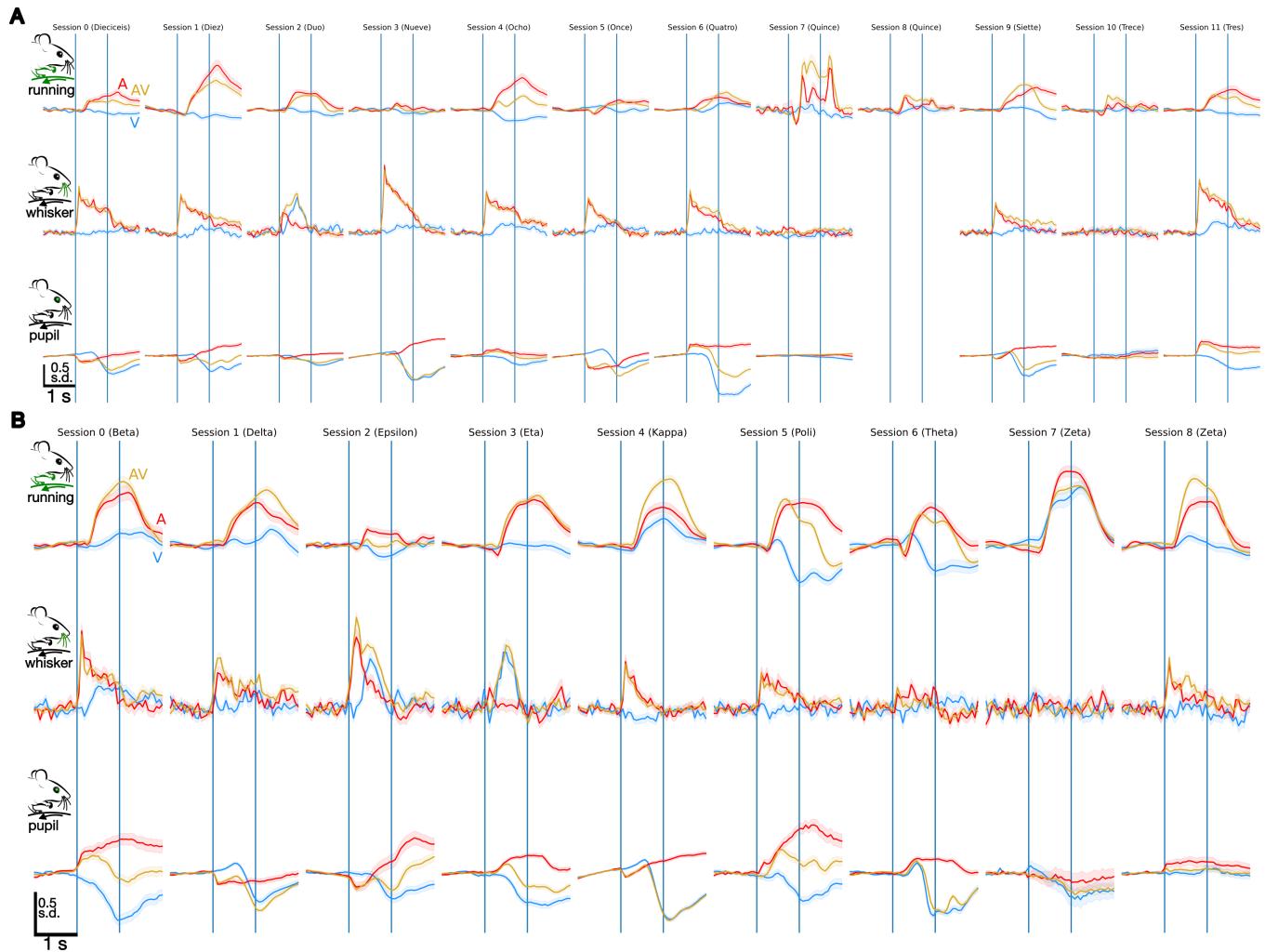
5 ACKNOWLEDGEMENTS

I would like to thank Huub Terra for supervision and support throughout this research project; Alexander Heimel and Leander de Kraker for helpful methodological discussions; and Christiaan Levert for hosting me at his lab.

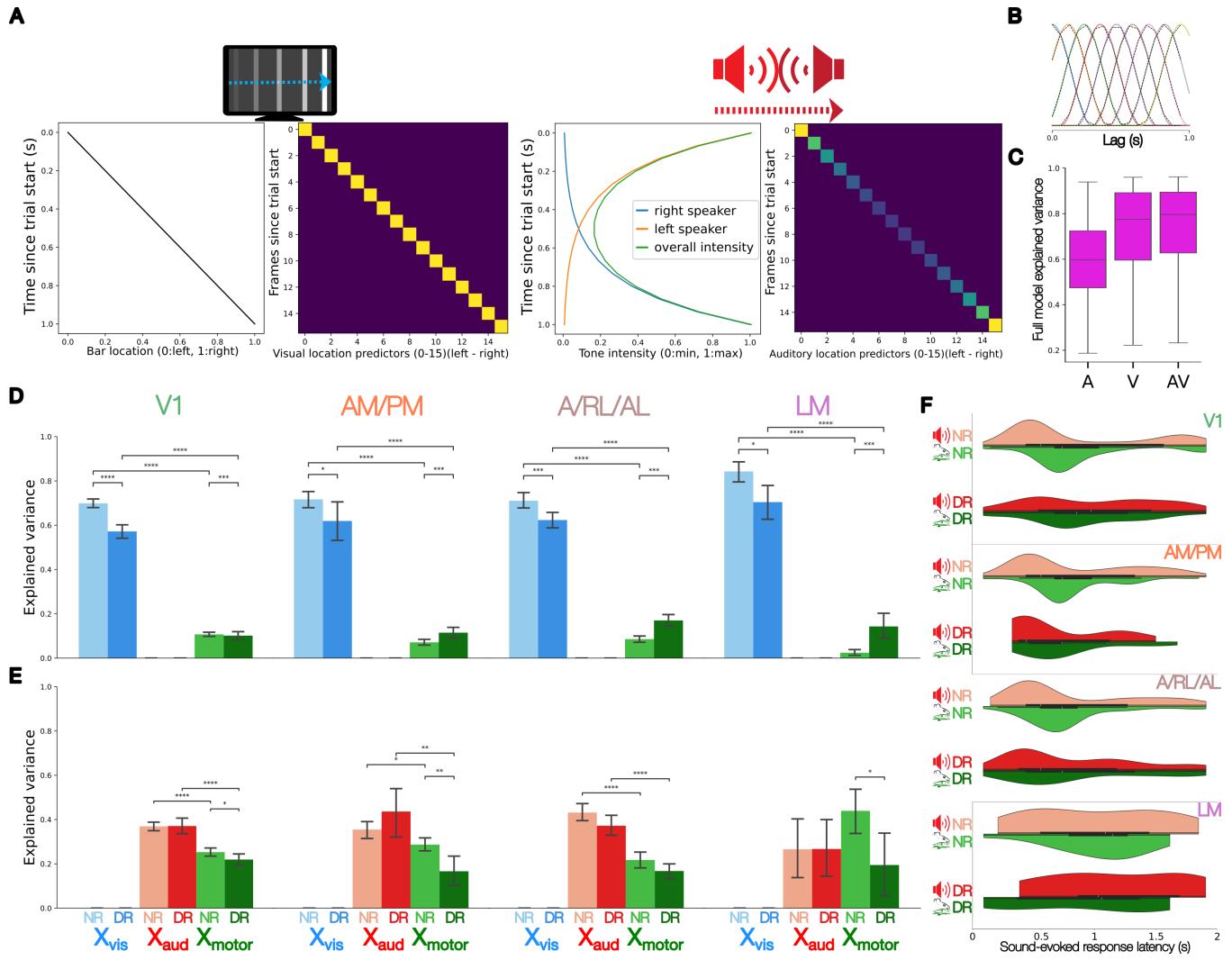
REFERENCES

- [1] Bavelier, D. & Neville, H. J. Cross-modal plasticity: where and how? *Nature Reviews Neuroscience* **3**, 443–452 (2002). MAG ID: 1557208597.
- [2] Ghazanfar, A. A. & Schroeder, C. E. Is neocortex essentially multisensory? *Trends in Cognitive Sciences* **10**, 278–285 (2006). MAG ID: 2073227393.
- [3] Fetsch, C. R., DeAngelis, G. C. & Angelaki, D. E. Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience* **14**, 429–442 (2013). MAG ID: 2043299443.
- [4] Stein, B. E., Stanford, T. R. & Rowland, B. A. Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience* **15**, 520–535 (2014). MAG ID: 2051387206.
- [5] Pennartz, C. M. A., Oude Lohuis, M. N. & Olcese, U. How 'visual' is the visual cortex? The interactions between the visual cortex and other sensory, motivational and motor systems as enabling factors for visual perception. *Philosophical Transactions of the Royal Society B: Biological Sciences* **378** (2023). URL <https://royalsocietypublishing.org/doi/10.1098/rstb.2022.0336>. Publisher: The Royal Society.
- [6] Röder, B., Kekunaya, R. & Guerreiro, M. J. S. Neural mechanisms of visual sensitive periods in humans. *Neuroscience & Biobehavioral Reviews* **120**, 86–99 (2020). MAG ID: 3108446410.
- [7] Hubel, D. H. & Wiesel, T. N. RECEPTIVE FIELDS OF CELLS IN STRIATE CORTEX OF VERY YOUNG, VISUALLY INEXPERIENCED KITTENS. *Journal of Neurophysiology* **26**, 994–1002 (1963). URL <https://www.physiology.org/doi/10.1152/jn.1963.26.6.994>.
- [8] Wiesel, T. N. & Hubel, D. H. SINGLE-CELL RESPONSES IN STRIATE CORTEX OF KITTENS DEPRIVED OF VISION IN ONE EYE. *Journal of Neurophysiology* **26**, 1003–1017 (1963). URL <https://www.physiology.org/doi/10.1152/jn.1963.26.6.1003>.
- [9] Putzar, L., Goerendt, I. K., Lange, K., Rösler, F. & Röder, B. Early visual deprivation impairs multisensory interactions in humans. *Nature Neuroscience* **10**, 1243–1245 (2007). MAG ID: 2023711861.
- [10] Putzar, L., Höttig, K. & Röder, B. Early visual deprivation affects the development of face recognition and of audio-visual speech perception. *Restorative Neurology and Neuroscience* **28**, 251–257 (2010). MAG ID: 1573570451.
- [11] Guerreiro, M. J. S., Putzar, L. & Röder, B. The effect of early visual deprivation on the neural bases of multisensory processing. *Brain* **138**, 1499–1504 (2015). MAG ID: 2097385647.
- [12] Collignon, O. *et al.* Long-Lasting Crossmodal Cortical Reorganization Triggered by Brief Postnatal Visual Deprivation. *Current Biology* **25**, 2379–2383 (2015). MAG ID: 1655458177.
- [13] Felleman, D. J. & Van Essen, D. C. Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex* **1**, 1–47 (1991). URL <https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/1.1.1>.
- [14] DSouza, R. D. *et al.* Hierarchical and nonhierarchical features of the mouse visual cortical network. *Nature Communications* **13**, 503 (2022). URL <https://www.nature.com/articles/s41467-022-28035-y>.
- [15] Braddick, O., Atkinson, J. & Wattam-Bell, J. Normal and anomalous development of visual motion processing: motion coherence and 'dorsal-stream vulnerability'. *Neuropsychologia* **41**, 1769–1784 (2003). URL <https://linkinghub.elsevier.com/retrieve/pii/S0028393203001787>.
- [16] Smith, I. T., Townsend, L. B., Huh, R., Zhu, H. & Smith, S. L. Stream-dependent development of higher visual cortical areas. *Nature Neuroscience* **20**, 200–208 (2017). URL <https://www.nature.com/articles/nn.4469>.
- [17] Olcese, U., Iurilli, G. & Medini, P. Cellular and Synaptic Architecture of Multisensory Integration in the Mouse Neocortex. *Neuron* **79**, 579–593 (2013). URL <https://linkinghub.elsevier.com/retrieve/pii/S0896627313005266>.
- [18] Meijer, G. T. *et al.* Neural Correlates of Multisensory Detection Behavior: Comparison of Primary and Higher-Order Visual Cortex. *Cell Reports* **31**, 107636 (2020). MAG ID: 3024943549 S2ID: af0b3af958c475a46b108ecee7eec96a257adb353.
- [19] Meijer, G. T., Montijn, J. S., Pennartz, C. M. A. & Lansink, C. S. Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency. *The Journal of Neuroscience* **37**, 8783–8796 (2017). MAG ID: 2743557698.
- [20] Knöpfel, T. *et al.* Audio-visual experience strengthens multisensory assemblies in adult mouse visual cortex. *Nature Communications* **10**, 5684–5684 (2019). MAG ID: 2995796775.
- [21] Ibrahim, L. *et al.* Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition. *Neuron* **89**, 1031–1045

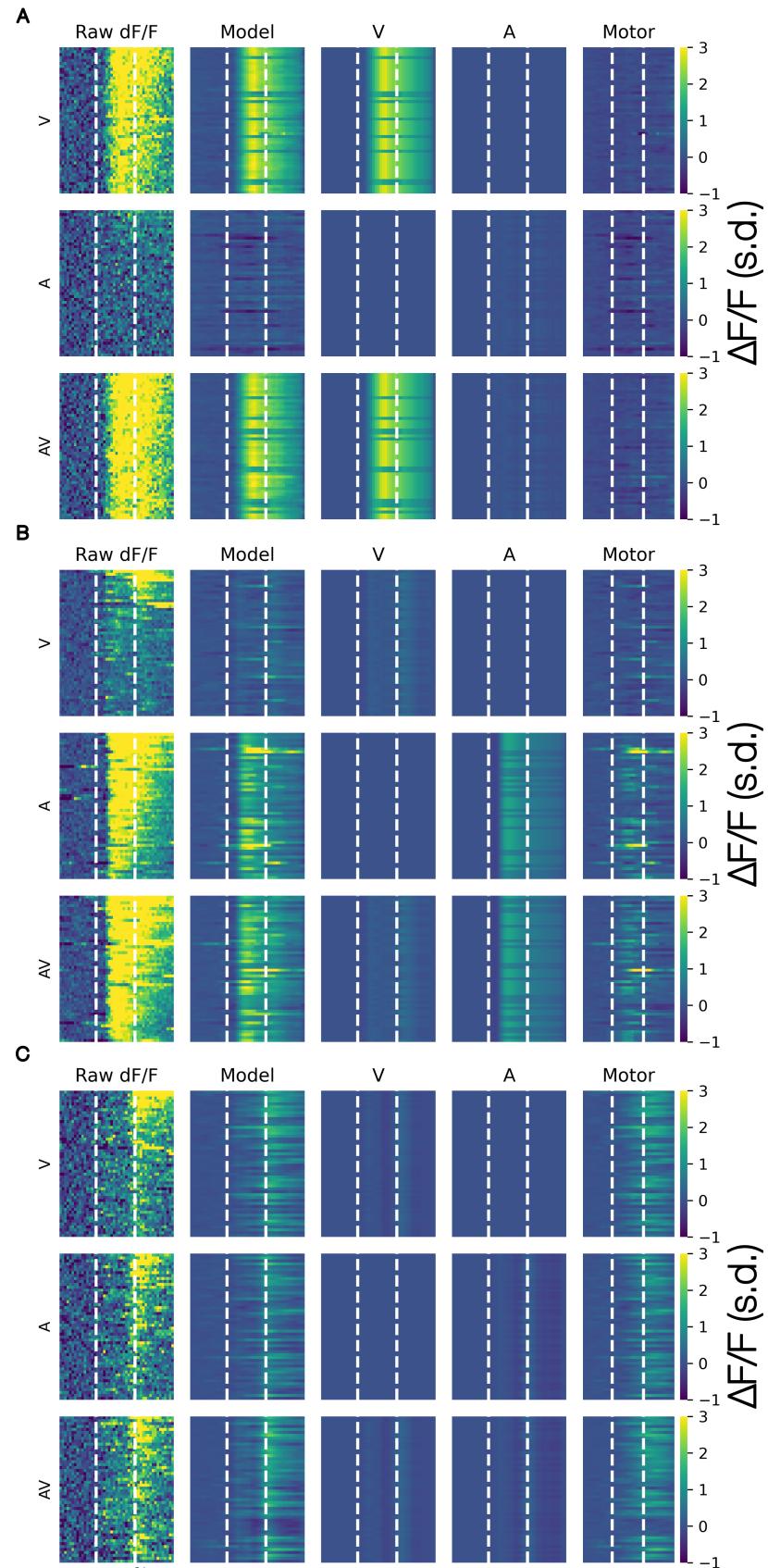
- (2016). URL <https://linkinghub.elsevier.com/retrieve/pii/S0896627316000520>. Publisher: Elsevier BV.
- [22] Williams, A. M., Angeloni, C. F. & Geffen, M. N. Sound Improves Neuronal Encoding of Visual Stimuli in Mouse Primary Visual Cortex. *The Journal of Neuroscience* **43**, 2885–2906 (2023). URL <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2444-21.2023>. Publisher: Society for Neuroscience.
- [23] Kriegstein, K. V., Kleinschmidt, A., Sterzer, P. & Giraud, A.-L. Interaction of Face and Voice Areas during Speaker Recognition. *Journal of Cognitive Neuroscience* **17**, 367–376 (2005). URL <https://direct.mit.edu/jocn/article/17/3/367/3996/Interaction-of-Face-and-Voice-Areas-during-Speaker>.
- [24] Renier, L. A. *et al.* Preserved Functional Specialization for Spatial Processing in the Middle Occipital Gyrus of the Early Blind. *Neuron* **68**, 138–148 (2010). URL <https://linkinghub.elsevier.com/retrieve/pii/S0896627310007634>.
- [25] Wallace, M. T., Ramachandran, R. & Stein, B. E. A revised view of sensory cortical parcellation. *Proceedings of the National Academy of Sciences* **101**, 2167–2172 (2004). URL <https://pnas.org/doi/full/10.1073/pnas.0305697101>. Publisher: Proceedings of the National Academy of Sciences.
- [26] Terra, H., De Kraker, L. & Levelt, C. Visual experience-dependent auditory and visual plasticity in the mouse visual cortex. *bioRxiv* (2024). S2ID: 983666852b0c0b4b4a7e70604ea8941ea55398b.
- [27] Ko, H. *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011). URL <https://www.nature.com/articles/nature09880>. Publisher: Springer Science and Business Media LLC.
- [28] Ko, H., Mrsic-Flogel, T. D. & Hofer, S. B. Emergence of Feature-Specific Connectivity in Cortical Microcircuits in the Absence of Visual Experience. *The Journal of Neuroscience* **34**, 9812–9816 (2014). URL <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0875-14.2014>. Publisher: Society for Neuroscience.
- [29] Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364** (2019). URL <https://www.science.org/doi/10.1126/science.aav7893>. Publisher: American Association for the Advancement of Science (AAAS).
- [30] Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience* **22**, 1677–1686 (2019). URL <https://www.nature.com/articles/s41593-019-0502-4>. Publisher: Springer Science and Business Media LLC.
- [31] Bimbard, C. *et al.* Behavioral origin of sound-evoked activity in mouse visual cortex. *Nature Neuroscience* **26**, 251–258 (2023). URL <https://www.nature.com/articles/s41593-022-01227-x>. Publisher: Springer Science and Business Media LLC.
- [32] Oude Lohuis, M. N., Marchesi, P., Olcese, U. & Pennartz, C. M. A. Triple dissociation of visual, auditory and motor processing in mouse primary visual cortex. *Nature Neuroscience* **27**, 758–771 (2024). URL <https://www.nature.com/articles/s41593-023-01564-5>. Publisher: Springer Science and Business Media LLC.
- [33] Galletti, C. & Fattori, P. The dorsal visual stream revisited: Stable circuits or dynamic pathways? *Cortex* **98**, 203–217 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0010945217300151>. Publisher: Elsevier BV.
- [34] Turner, K. L., Gheres, K. W. & Drew, P. J. Relating Pupil Diameter and Blinking to Cortical Activity and Hemodynamics across Arousal States. *The Journal of Neuroscience* **43**, 949–964 (2023). URL <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1244-22.2022>. Publisher: Society for Neuroscience.
- [35] Fitzpatrick, M. J., Krizan, J., Hsiang, J.-C., Shen, N. & Kerschensteiner, D. A pupillary contrast response in mice and humans: Neural mechanisms and visual functions. *Neuron* **112**, 2404–2422.e9 (2024). URL <https://linkinghub.elsevier.com/retrieve/pii/S0896627324002733>. Publisher: Elsevier BV.
- [36] Akella, S. *et al.* Deciphering neuronal variability across states reveals dynamic sensory encoding. *Nature Communications* **16** (2025). URL <https://www.nature.com/articles/s41467-025-56733-w>. Publisher: Springer Science and Business Media LLC.
- [37] Karlen, S., Kahn, D. & Krubitzer, L. Early blindness results in abnormal cortico-cortical and thalamocortical connections. *Neuroscience* **142**, 843–858 (2006). URL <https://linkinghub.elsevier.com/retrieve/pii/S030645220600861X>. Publisher: Elsevier BV.
- [38] De Heering, A. *et al.* A Brief Period of Postnatal Visual Deprivation Alters the Balance between Auditory and Visual Attention. *Current Biology* **26**, 3101–3105 (2016). URL <https://linkinghub.elsevier.com/retrieve/pii/S096098221631199X>. Publisher: Elsevier BV.
- [39] Mazo, C., Baeta, M. & Petreanu, L. Auditory cortex conveys non-topographic sound localization signals to visual cortex. *Nature Communications* **15** (2024). URL <https://www.nature.com/articles/s41467-024-47546-4>. Publisher: Springer Science and Business Media LLC.
- [40] Charbonneau, V., Laramée, M., Boucher, V., Bronchti, G. & Boire, D. Cortical and subcortical projections to primary visual cortex in anophthalmic, enucleated and sighted mice. *European Journal of Neuroscience* **36**, 2949–2963 (2012). URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1460-9568.2012.08215.x>.
- [41] Wallace, M. S. *et al.* Visual experience is necessary for the development of multisensory integration. *The Journal of Neuroscience* **24**, 9580–9584 (2004). MAG ID: 2092425177.
- [42] Wallace, M. T. & Stein, B. E. Early experience determines how the senses will interact. *Journal of Neurophysiology* **97**, 921–926 (2007). MAG ID: 2148235578.
- [43] Smyre, S. A., Wang, Z., Stein, B. E. & Rowland, B. A. Multisensory enhancement of overt behavior requires multisensory experience. *European Journal of Neuroscience* **54**, 4514–4527 (2021). URL <https://onlinelibrary.wiley.com/doi/10.1111/ejn.15315>. Publisher: Wiley.
- [44] Stanford, T. R., Quesy, S. & Stein, B. E. Evaluating the Operations Underlying Multisensory Integration in the Cat Superior Colliculus. *The Journal of Neuroscience* **25**, 6499–6508 (2005). URL <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.5095-04.2005>. Publisher: Society for Neuroscience.
- [45] Iurilli, G. *et al.* Sound-Driven Synaptic Inhibition in Primary Visual Cortex. *Neuron* **73**, 814–828 (2012). MAG ID: 213099752.
- [46] Carlini, A., Bordeau, C. & Ambard, M. Auditory localization: a comprehensive practical review. *Frontiers in Psychology* **15**, 1408073 (2024). URL <https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1408073/full>.
- [47] Ito, S., Si, Y., Feldheim, D. A. & Litke, A. M. Spectral cues are necessary to encode azimuthal auditory space in the mouse superior colliculus. *Nature Communications* **11**, 1087 (2020). URL <https://www.nature.com/articles/s41467-020-14897-7>.
- [48] Ito, S., Si, Y., Litke, A. M. & Feldheim, D. A. Nonlinear visuoauditory integration in the mouse superior colliculus. *PLOS Computational Biology* **17**, e1009181 (2021). URL <https://dx.plos.org/10.1371/journal.pcbi.1009181>.
- [49] Garrett, M. E., Nauhaus, I., Marshel, J. H. & Callaway, E. M. Topography and Areal Organization of Mouse Visual Cortex. *The Journal of Neuroscience* **34**, 12587–12600 (2014). URL <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1124-14.2014>. Publisher: Society for Neuroscience.
- [50] Van Beest, E. H. *et al.* Mouse visual cortex contains a region of enhanced spatial resolution. *Nature Communications* **12** (2021). URL <https://www.nature.com/articles/s41467-021-24311-5>. Publisher: Springer Science and Business Media LLC.
- [51] de Kraker, L. *et al.* SpecSeg is a versatile toolbox that segments neurons and neurites in chronic calcium imaging datasets based on low-frequency cross-spectral power. *Cell reports methods* **100299**–100299 (2022). MAG ID: 4297902633 S2ID: 3e1f210a74c98e6844b44de89f5e62266f2dfa3d.
- [52] Pnevmatikakis, E. A. & Giovannucci, A. NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *Journal of Neuroscience Methods* **291**, 83–94 (2017). URL <https://linkinghub.elsevier.com/retrieve/pii/S0165027017302753>. Publisher: Elsevier BV.
- [53] Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* **181**, 936–953.e20 (2020). URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867420304025>. Publisher: Elsevier BV.
- [54] Syeda, A. *et al.* Facemap: a framework for modeling neural activity based on orofacial tracking. *Nature Neuroscience* **27**, 187–195 (2024). URL <https://www.nature.com/articles/s41593-023-01490-6>. Publisher: Springer Science and Business Media LLC.
- [55] Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*.
- [56] Berens, P. Comparing the feature selectivity of the gamma-band of the local field potential and the underlying spiking activity in primate visual cortex. *Frontiers in Systems Neuroscience* **2** (2008). URL <http://journal.frontiersin.org/article/10.3389/neuro.06.002.2008/abstract>.



Extended Data Figure 1: Session-averaged movement signals across sessions, trial types and rearing conditions. **A**) Session data for normally reared group (12 sessions, 11 mice). **B**) Session data for dark-reared group (9 sessions, 8 mice). (**A, B**) Running (top), whisking (middle), pupil size (bottom). Columns represent separate sessions. Visual trials (blue), auditory trials (red) and audiovisual trials (gold). Shaded regions indicate ± 1 SEM.

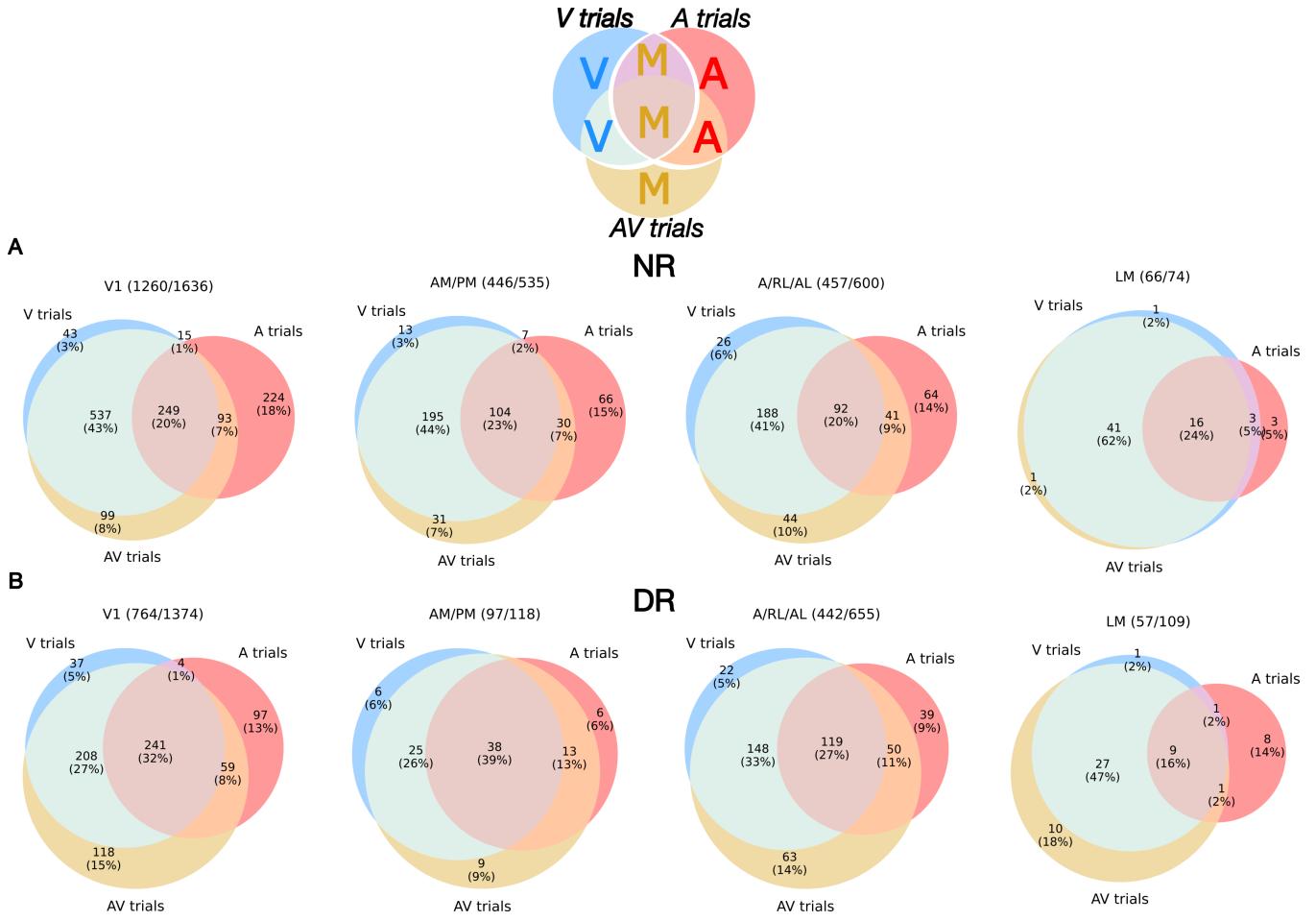


Extended Data Figure 2: Regression model design, explained variance and predicted response latency in unimodal trials. **A)** Stimulus modeling in an example audiovisual trial (ArVr). Example rightward visual stimulus (Vr), and example rightward auditory stimulus (Ar). Representation of example rightward stimuli in the design matrix X of the regression model with separate "location" predictors as columns (right). **B)** Example raised cosine bases used to model different temporal lags between predictors and neural responses. Nine bases spanning lags between 0-1 s are shown. Dashed lines indicate raised cosine bases sampled at the 2P imaging resolution (≈ 15.5 Hz). **C)** Explained variance of trial-averaged single-neuron activity by the full model across trial types. Neurons with significant explained variance by the full model for 20% held-out trials of any type were used (NR: 2266 neurons; DR: 1349 neurons). **D)** Same as Fig 3C, but for visual trials. Neurons with significant explained variance by the full model for held-out visual trials were used (NR: 1008 neurons, DR: 560 neurons). **E)** Same as Fig 3C, but for auditory trials. **F)** Distribution of single-neuron sound-evoked response latencies (auditory trials) predicted by auditory versus motor predictors across brain areas and rearing conditions. Midline of each violin plot contains box plots summarizing the distribution. **(E, F)** Neurons with significant explained variance by the full model for held-out auditory trials were used (NR: 747 neurons, DR: 378 neurons). **(D, E)** Explained variance was calculated on 20% held-out trials. Error bars indicate 95% confidence intervals; between-group comparisons (NR vs DR) were done using Mann-Whitney U test with Bonferroni correction; and within-group comparisons between different predictor sets were done using Wilcoxon Signed-Rank test with Bonferroni correction. (* : $p \leq .05$; ** : $p \leq .01$; *** : $p \leq .001$, *** : $p \leq .0001$)

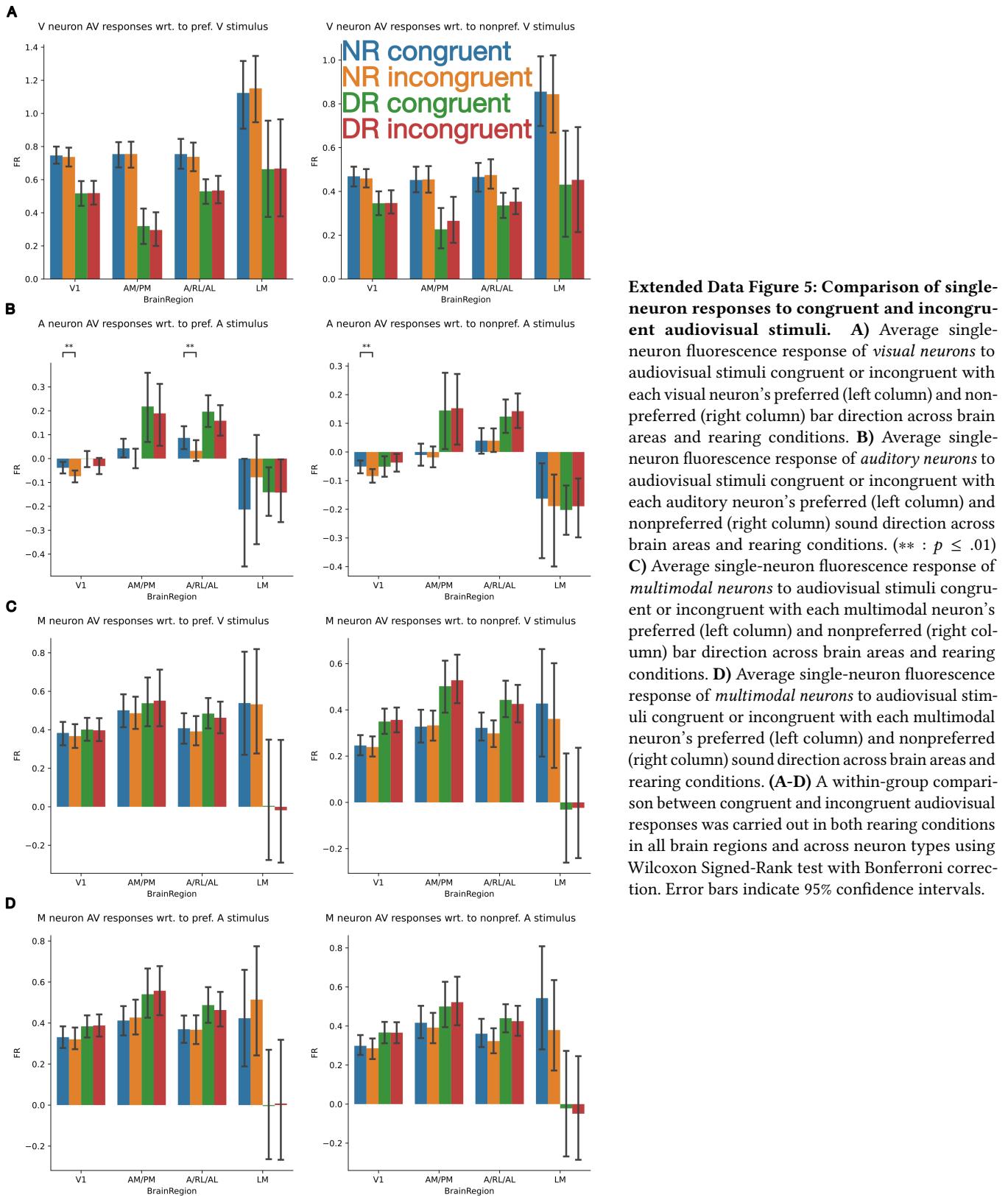


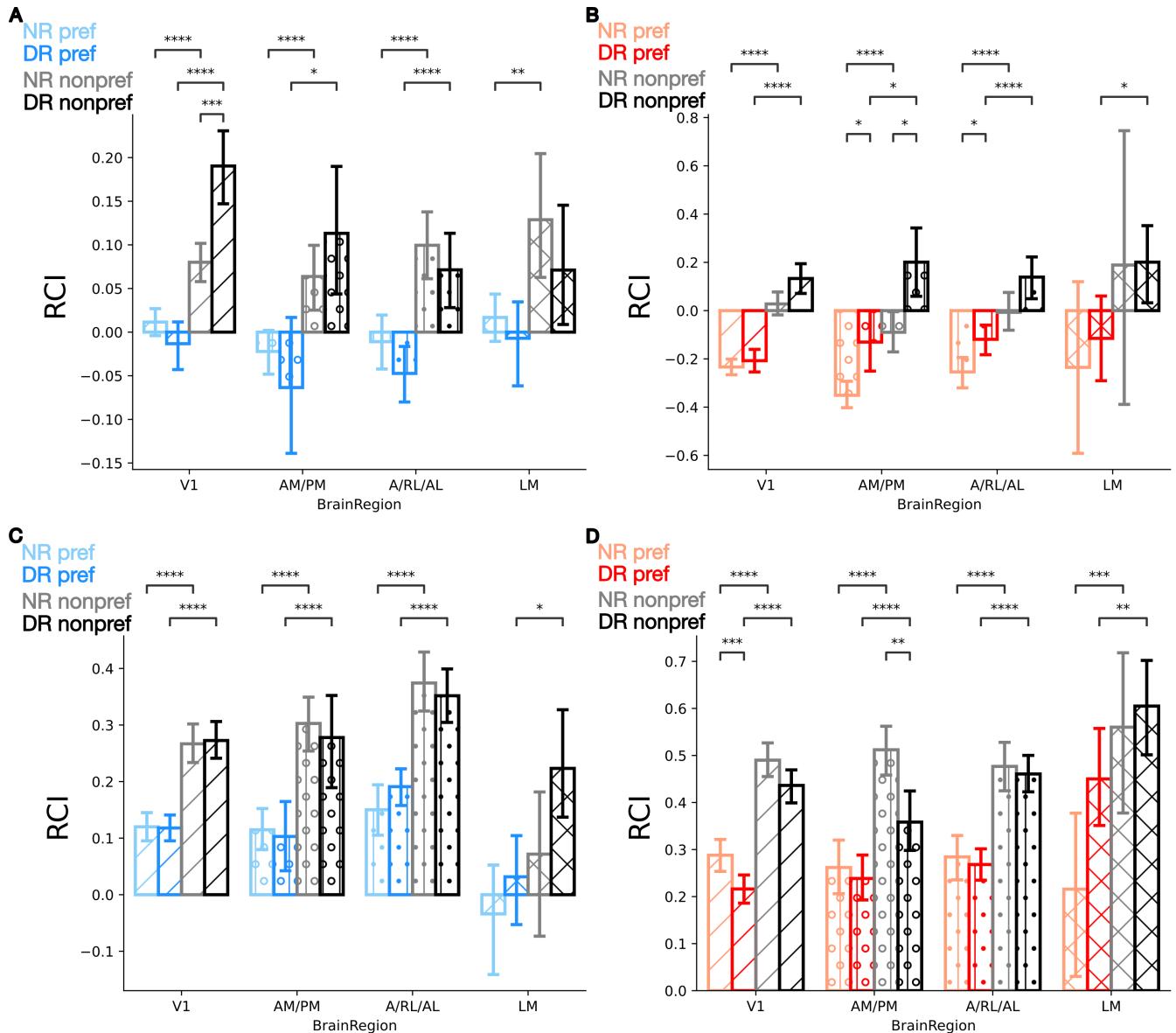
Extended Data Figure 3: Dissociation of sensory and motor-related neural activity in example visual cortex neurons using a regression model.

Each heatmap shows z-scored single-neuron $\Delta F/F$ activity for 50 % of trials with the highest original $\Delta F/F$ activity in a given trial type (V, A, AV - rows). First column shows the neuron's raw, originally measured $\Delta F/F$ responses. Second column shows neural activity predicted by the full regression model including the additive contribution from all predictor sets. Columns three to five show neural activity predicted by visual, auditory and motor predictor subsets respectively. **A)** Example LM neuron (Neuron ID: 38) with strong vision-related activity and little auditory or motor-related activity. **B)** Example V1 neuron (Neuron ID: 429) with strong auditory-related activity, occasional bursts of motor-related activity and little vision-related activity. **C)** Example A/RL/AL neuron (Neuron ID: 415) with dominant motor-related activity and virtually absent sensory-related activity.



Extended Data Figure 4: Trial-type responsiveness criteria and statistics for functional neuron type grouping. Top center: Illustration of trial-type responsiveness inclusion criteria for each neuron type V: unimodal visual neurons, A: unimodal auditory neurons, M: multimodal neurons. **A)** Single-neuron trial-type responsiveness statistics across brain areas for normally reared group (12 sessions, 11 mice). **B)** Single-neuron trial-type responsiveness statistics across brain areas for dark-reared group (9 sessions, 8 mice). (**A-B**) Fraction next to area name indicates count of stimulus responsive neurons out of all recorded neurons in that area. Within each subset of each Venn diagram, count of stimulus-responsive neurons in that subset is shown, as well as, what proportion of stimulus-responsive neurons from that area the subset represents.





Extended Data Figure 6: Single-neuron multisensory integration across visual cortex of dark-reared and normally reared mice. **A)** Unimodal visual neurons, same as Fig. 5D, but calculated with RCI instead of Absolute RCI. **B)** Unimodal auditory neurons, same as Fig. 5H, but calculated with RCI instead of Absolute RCI. **C)** Multimodal neurons, same as Fig. 5L, but calculated with RCI instead of Absolute RCI. **D)** Multimodal neurons, same as Fig. 5M, but calculated with RCI instead of Absolute RCI. **(A-D)** Error bars indicate 95% confidence intervals; between-group comparisons (NR vs DR) were done using Mann-Whitney U test with Bonferroni correction; and within-group comparisons between preferred and nonpreferred directions were done using Wilcoxon Signed-Rank test with Bonferroni correction. (* : $p \leq .05$; ** : $p \leq .01$; *** : $p \leq .001$, **** : $p \leq .0001$)