Different representations in layer 2/3 and layer 5 excitatory neurons of the primary visual cortex

Timothy P.H. Sit,[a,b,*] Célian Bimbard,[c] Anna Lebedeva,[a,b] Matteo Carandini,[c] Philip Coen,[c,d] Kenneth D Harris[b]

[a]Sainsbury Wellcome Centre, University College London, London, UK.
[b]UCL Queen Square Institute of Neurology, University College London, London, UK
[c]UCL Institute of Ophthalmology, University College London, London, UK
[d]Department of Cell and Developmental Biology, University College London, London, UK
*Correspondence should be addressed to T.P.H.S. or K.D.H. (sitpakhang@gmail.com, kenneth.harris@ucl.ac.uk)

The cortex contains multiple types of excitatory neuron, differentiated primarily by their layer of residence. We recorded from neuronal populations in the mouse visual cortex using 2-photon calcium imaging or Neuropixels probes, and found that excitatory neurons in layer 2/3 (L2/3) and layer 5 (L5) differ in their encoding of visual vs. nonvisual signals, with L2/3 more strongly modulated by visual stimuli and L5 more strongly modulated by movement. Movement has opposite effects on population synchrony in the two layers, desynchronizing L2/3 by abolishing spontaneous population oscillations, and synchronizing L5, where excitatory cells are less entrained by oscillation and more strongly correlated with movement itself. Spontaneous activity is lower-dimensional in L2/3 than L5, with L2/3 population activity dominated by a single dimension of overlap between spontaneous and stimulus-evoked subspaces. We conclude that the population code of in different layers of the visual cortex differentially balances visual and non-visual signals.

Introduction

The cortex contains multiple types of excitatory neuron, which are primarily distinguished by the layer in which their soma is located. Excitatory populations of different layers receive different sets of synaptic inputs, have different intrinsic physiological properties, and send their axonal outputs to different targets[1–8] . They also use different strategies to encode information[9–15]. Characterizing these strategies is an essential step towards understanding the role of these different populations in cortical information processing.

Traditionally, neural codes have been distinguished at a single-cell level, for example by contrasting the types of sensory stimuli that maximally drive cells of different populations. Now it is possible to record from large populations simultaneously, allowing us to understand the structure of population coding directly[16]. The way that these populations encode information is not necessarily static: the cortex can operate in multiple states, depending on the animal's level of alertness[17–20], and the structure of the population code might change between states. Furthermore, even in primary sensory cortex, activity is not fully determined by sensory inputs, but shows highly structured multidimensional spontaneous activity, partly determined by an animal's spontaneous movements[21,22].

Metrics to characterize the format of population codes are still evolving, but some simple strategies are emerging. One can characterize a neuron's relationship to the population by its "population coupling" coefficient - the correlation of its activity with the population rate (the summed activity of all recorded cells)[23]. The distribution of population couplings across neurons summarizes the degree of synchrony in the population: when most neurons have high population coupling it reflects a synchronized state, where neuronal activity fluctuates together synchronously; when population coupling is lower, it represents a state where positive or negative correlations between neurons may exist, but do not cause common fluctuations in summed activity[24,25].

A second way to characterize population activity is by geometry and dimensionality. One can consider the activity of N neurons as defining a vector in an N-dimensional space. The activity vectors produced by a brain circuit are typically constrained to a small part of this space, such as a subspace or manifold. Although characterizing the nonlinear structure of activity manifolds is challenging[26], one can characterize the linear spaces that activity is constrained to by variants of principal components analysis, which measure the sizes of dimensions occupied by activity vectors. Because cortical activity reflects both sensory responses and spontaneous, behavior-related activity, it is important to separately characterize the geometry of the subspaces defined by spontaneous activity and sensory responses, and to understand the degree to which these subspaces overlap. We have previously defined methods to perform these characterizations[21,27], but only applied them to L2/3 populations.

Here we contrast the way that neuronal populations of superficial and deep visual cortex encode movement and visual signals, using 2-photon calcium imaging and electrophysiology. We find that activity in the deep layers more strongly encodes movement, while activity in the superficial layers more strongly encodes visual stimuli. The geometry of the population code differed between layers, with activity in superficial layers more strongly dominated by a single dimension, which was particularly prominent for superficial-layer spontaneous activity in stationary states, where spontaneous activity became highly synchronized along this single dimension.

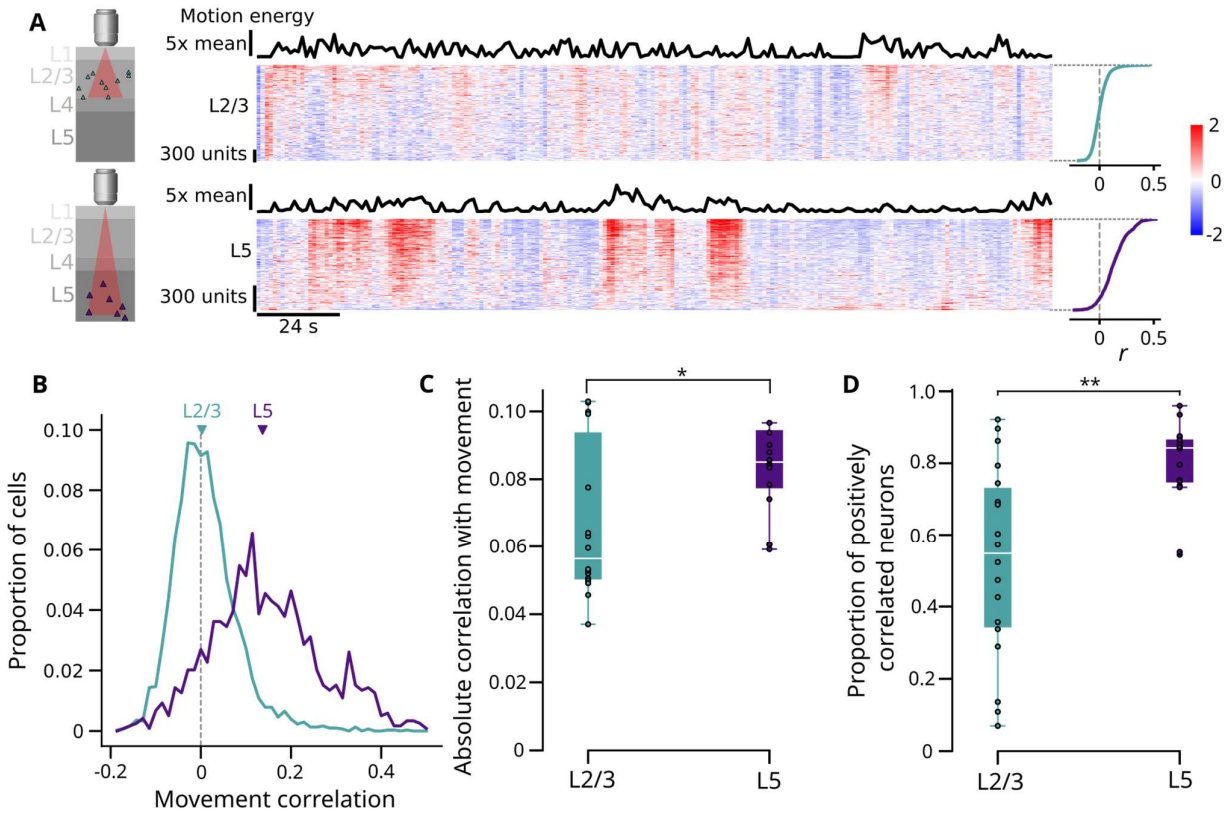

Figure 1 Movement correlates are more uniform in L5 than L2/3. (A) Top: example 2-photon recording of L2/3 spontaneous activity, with videographic motion energy shown as a black curve (sub-sampled to 2.5 Hz), and L2/3 population activity shown as a pseudocolor raster with cells vertically arranged by movement correlation (2.5 Hz sampling rate). Right: a curve showing the Pearson correlation of neural activity with movement energy (x-axis) for cells ordered as in the raster (y-axis). Bottom: same for L5 neurons. **(B)** Histograms of movement correlations of all cells in L2/3 (green) and L5 (purple) example experiments shown in (A). Triangles show mean movement correlation values of the example recording sessions. **(C)** Mean absolute correlation of neurons with movement, averaged over all cells in each session (one dot per session), in L2/3 and L5 ($p = 0.02$, rank sum test, N = 18 and 16 sessions in L2/3 and L5). **(D)** Proportion of cells in each session positively correlated with movement ($p = 0.003$), rank sum test, N = 18 and 16 sessions in L2/3 and L5). * $p < 0.05$, ** $p < 0.01$.

Results

To investigate the differences in the encoding of sensory-evoked and spontaneous activity between cortical layers, we performed two-photon calcium imaging of visual cortical populations in two mouse lines. To record from layer 2/3 (L2/3) excitatory neurons we used the TetO-GCaMP6s;CaMKIIa-tTA mouse line[28], which labels all excitatory cells, resulting in good recordings of superficial layers but occlusion of deep layers. To record from deeper layers, we crossed Rbp4-Cre KL100[29,30] with Ai94 (TITL-GCaMP6s)[31], which labels layer 5 (L5) excitatory neurons without occlusion because more superficial cells are not labelled. In each session, we recorded spontaneous activity (gray-screen conditions) and responses to combinations of natural videos and sounds, as previously described[32]. We monitored the orofacial movement of mice videographically during the experiments. In addition to recording large populations with 2-photon microscopy, we analyzed the responses of smaller populations recorded with Neuropixels probes in response to the same stimuli.

Movement correlates are more uniform in L5 than L2/3

The correlation of V1 activity with movement was stronger and more positive in L5 than L2/3. We computed each neuron's Pearson's correlation with videographic motion energy during the spontaneous period, when only a gray background is presented on the three screens. Consistent with previous results[21], in L2/3 we observed both positive and negative correlations with movement energy. By contrast, in L5 correlations were larger and more commonly positive (Figure 1A, B). The negative correlations with movement in L2/3 were genuine: a linear shift test[33] revealed that of L2/3 cells that significantly correlated with movement, 61% ± 9% correlated positively and the remainder negatively. However in L5 almost all significant correlations were positive (97% ± 2%; Figure S1). Correlations (regardless of significance) were stronger in L5 (i.e. larger absolute value; Fig. 1C; $p = 0.02$, Rank sum test, N = 18 and 16 sessions in L2/3 and L5) and a larger fraction of correlations (whether significant or not) were positive in L5 than L2/3 (Figure 1D; $p = 0.003$, Rank sum test, N = 18 and 16 sessions in L2/3 and L5). Analysis of Neuropixels recordings in V1[32] showed similar results (Figure S2A-D).

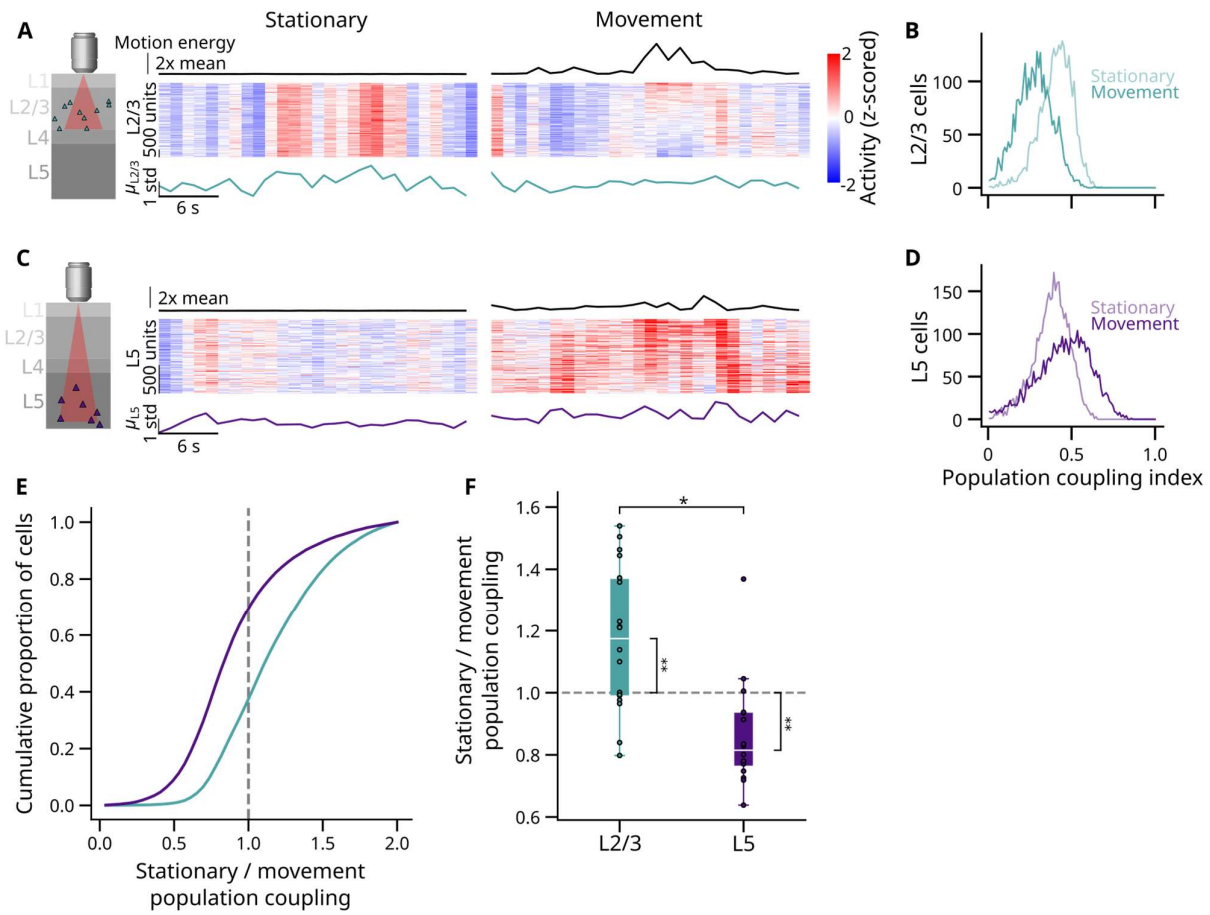

Figure 2. Movement desynchronizes L2/3 but synchronizes L5. (A) Pseudocolor rasters of L2/3 population activity during periods of stationarity (left) or movement (right), with neurons arranged vertically by correlation with movement as in Figure 1A. Above: videographic motion energy. Below, population rate, i.e. summed activity of all recorded neurons. **(B)** Histogram of population coupling indices (i.e. correlation of each neuron with population rate) for L2/3 neurons during movement and stationary periods. **(C-D)** Same as (A-B), for L5 neurons. **(E)** Cumulative histogram of the ratio of population coupling in stationary vs. movement periods, for L2/3 (green) and L5 (purple) neurons. **(F)** Comparison of the ratio of stationary and movement population coupling indices in L2/3 and L5 neurons. Each dot represents the mean coupling of all neurons in one recording session ($p = 0.01$, linear mixed effects model, N = 5 mice for both L2/3 and L5, N = 18 sessions for L2/3 and 16 sessions for L5). * $p < 0.05$, ** $p < 0.01$.

Although the encoding of movement in population activity had different forms in L2/3 and L5, population activity in both layers could predict movement. We used linear regression to predict motion energy from the first 40 principal components of neural activity acquired using two-photon calcium imaging (Figure S3A, B). We could significantly predict motion energy above chance from activity in either L2/3 and L5 (Figure S3C-E), with no significant difference in prediction quality between layers (Figure S3F).

Movement desynchronizes L2/3 but synchronizes L5

Movement decreased spontaneous population synchrony in L2/3 but increased it in L5. We assessed population synchrony by computing each neuron's population coupling index[23], defined as the Pearson correlation of its activity with the population rate (i.e. the summed activity of all recorded neurons). L2/3 excitatory neurons exhibited strongly synchronous spontaneous oscillations when the mice were stationary (Fig. 2A), which reduced when the animal moved, leading to a decrease in population coupling (Fig. 2B). By contrast, L5 spontaneous oscillations were weaker during stationary periods and correlations with running were stronger, leading to an increase in population coupling during movement (Figure 2C, D). This difference was consistent across sessions: for each cell we computed the ratio of population coupling coefficients between stationary and movement states (Figure 2E), and found that this ratio was significantly greater than 1 in L2/3 populations but significantly < 1 in L5 populations, with a significant difference between layers (Figure 2F; $p = 0.01$, Linear mixed effects model; L2/3: N = 5 mice, 16 sessions; L5: N=5 mice, 16 sessions) Similar results were found with electrophysiology recordings (Figure S2E, F).

Spontaneous activity is lower-dimensional in L2/3 than L5

The strong population coupling seen in L2/3 spontaneous activity during stationarity suggests that the corresponding population activity may be dominated by a single dimension to which many neurons couple. To investigate this, we used

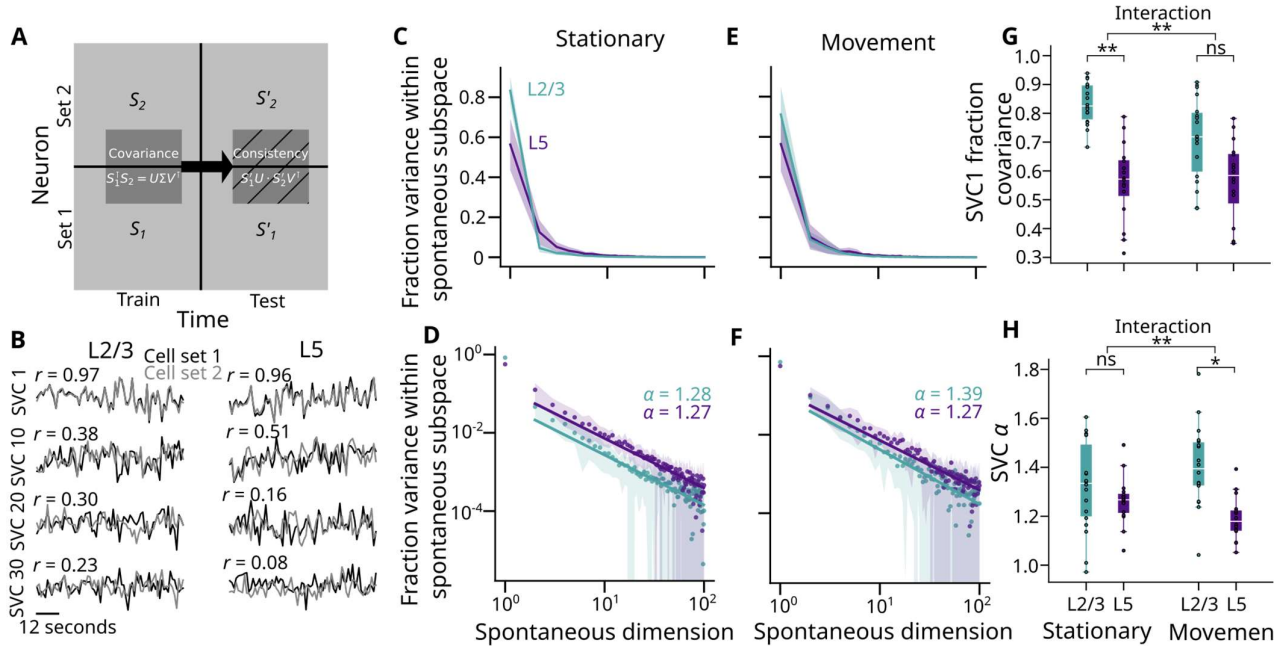

Figure 3.Spontaneous activity is lower-dimensional in L2/3 than L5. (A) Schematic of shared variance component analysis (SVCA). Neurons were randomly divided into two sets, and their covariance matrix was computed from training time points. The consistency of this shared activity was evaluated using testing time points. **(B)** Time traces of shared variance component projections in one example recording in L2/3 (left) and layer 5 (right), for 84 seconds of test-set activity. R-values represent the Pearson correlation between the shared variance component of cell set 1 and cell set 2 during testing time points. **(C)** Fraction of variance explained by each shared variance component in L2/3 (green) and L5 (purple) during stationary periods. Shading: standard error of the mean across mice. **(D)** As in (C), but with fraction variance in log-scale, and line of best fit of the second component onwards. **(E, F)** As in (C, D), but during movement periods. **(G)** Comparison of the fraction of variance explained by the first shared variance component in L2/3 and L5 populations during movement and stationary periods ($p = 0.003$ during stationary period, $p = 0.15$ during movement, interaction $p = 0.003$, linear mixed effects model. N = 5 mice for both L2/3 and L5, N = 18 sessions for L2/3 and 16 sessions for L5). **(H)** As in (G), but for the comparison of the variance-decay slope from the second component onwards ($p = 0.570$ during stationary period, $p = 0.0212$ during movement, interaction $p = 0.001$, linear mixed effects model). * $p < 0.05$, ** $p < 0.01$

shared variance component analysis (SVCA)[21], which identifies activity dimensions which are reliably shared across cells and time, and quantifies the amount of neural activity reliably found in each dimension (Figure 3A, B).

SVCA analysis of an example session confirmed that spontaneous L2/3 population activity is more dominated by a single dimension than L5, particularly during stationary periods (for stationary periods, first dimension fraction variance for L2/3: 83% ± 2%, L5: 56% ± 3%, for movement periods: L2/3: 71% ± 3%, L5: 56% ± 3%) (Figure 3C). Viewing the fraction of variance accounted for by each dimension on a log-log scale showed that the reliable variance in all but the first few dimensions of spontaneous activity decays as a power law with exponent ~1.3 (Figure 3D), as previously observed[21]. However, the first dimension had substantially higher variance than predicted by this power law, particularly in L2/3. During movement periods, the dominance of L2/3 activity by a single dimension was substantially reduced (Figure 3E- F). The first dimension accounted for significantly more variance in L2/3 than L5 during stationary, but not movement periods, and the interaction between layer and movement condition was significant (Figure 3G; Linear mixed effects model interaction term $p = 0.003$; N=34 sessions). The distribution of variance across higher dimensions was less affected by movement, although the exponent of the power law was larger in L2/3 than L5, specifically during movement periods (Figure 3H). This suggests that L2/3 spontaneous activity is lower-dimensional than L5, and that the nature of this difference depends on state: during stationary states it is dominated by a single dimension, while in movement periods the power-law exponent is larger.

Correlates of sound-evoked movements are larger in L5 than L2/3

We next investigated V1 responses to auditory stimuli in the different layers. Strikingly, we found a substantial discrepancy between our 2-photon data and previously-acquired Neuropixels recordings[32]: auditory-evoked responses were strong when with recorded Neuropixels, but not 2p (Figure S4A-G). Because V1 auditory responses correlate with sound-evoked movements [32], we hypothesized that the reason they do not occur in 2p recordings was that the background noise produced by the 2p microscope (not acoustically shielded in our setup) inhibited the auditory startle response normally evoked by a sound presentation. Consistent with this hypothesis, the same sounds evoked much weaker movements during two-photon calcium imaging than electrode recordings (Figure S4A, B, D). To determine if the background noise could explain this difference, we presented the same sounds to head-fixed mice in an acoustically isolated test rig, and in some of these experiments additionally played simulated 2p background noise (recorded during 2p experiments). Sound-evoked movements were significantly reduced when background noise was present, recapitulating

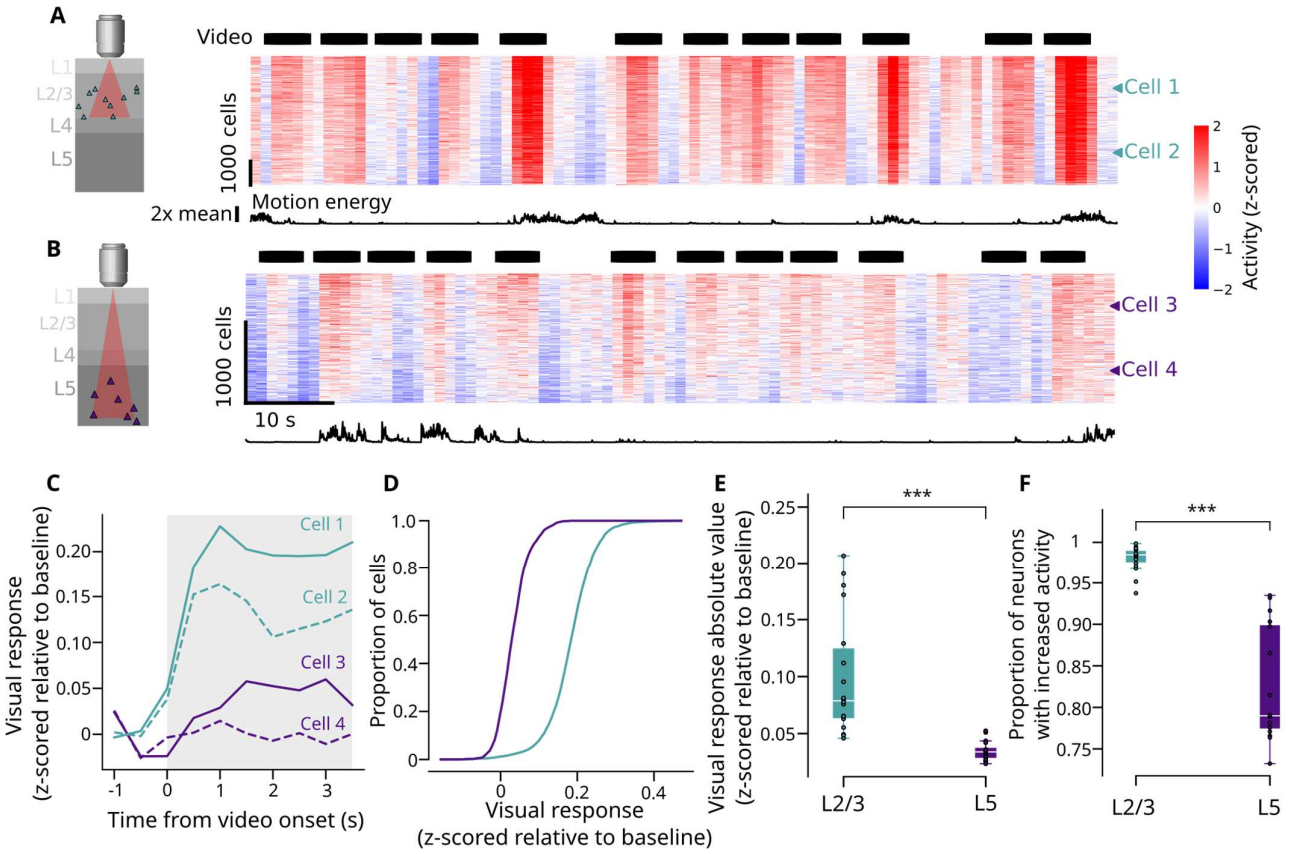

Figure 4. Visual onset responses are stronger in L2/3 than L5. (A) Pseudocolor activity raster for an example recording in L2/3. Black bars: times of video playback. **(B)** Same as (A) but for an example L5 recording. **(C)** Mean visual response across videos aligned to video onset (Z-scored relative to baseline) for four example cells indicated by arrows in (A) and (B). The gray region indicates the time window in which activity is averaged to compute video response shown in (D-F) **(D)** Histogram of mean visual response for cells in the example L2/3 (green) and L5 (purple) recordings. **(E)** Comparison of the mean absolute visual response between L2/3 and L5 neurons. Each dot represents the mean across neurons in a particular recording session ($p < 0.001$, rank sum test, N = 18 and 16 sessions in L2/3 and L5). **(F)** Same as (E), but for the proportion of neurons with increased activity after video onset ($p < 0.001$, rank sum test, N = 18 and 16 sessions in L2/3 and L5). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the difference between the Neuropixels and 2p recordings (Figure S4C-E). Consistent with the hypothesis that V1 auditory responses primarily reflect sound-evoked movements[32], sound-evoked V1 activity was much stronger in the electrode recordings than the 2p recordings, in either layer (Figure S4F-I).

We therefore quantified the difference in sound-evoked activity between layers only from our electrophsyiological recordings, which revealed stronger sound-evoked responses in L5 (Figure S4F, H), consitent with our earlier result showing stronger movement correlates in layer 5 (Figure 1). In summary, these data are consistent with V1 auditory responses being primarily dominated by sound-evoked movements, which lead to larger neuronal responses in L5 as this layer has stronger movement correlates. Visual onset responses are stronger in L2/3 than in L5

In contrast to movement correlates (which were stronger and more positive in L5), visual stimulus responses were stronger and more positive in L2/3 than L5. We computed each cell's mean response to natural videos, relative to pre-stimulus baseline (Figure 4A-C; see methods), and found responses to be stronger in L2/3 compared to L5 (Figure 4D), both in terms of the mean absolute value of response (Figure 4E) and the proportion of neurons with increased activity after visual onset (Figure 4F). Analysis of electrophysiology experiments produced consistent results (Figure S5A-C).

Population activity is more 1-dimensional in L2/3 than in L5

To estimate the amount of stimulus-related power in successive dimensions we used cross-validated principal components analysis (cvPCA)[27] (Figure 5A). This analysis was performed on data from movement and stationary states together, as it requires repeats of each stimulus which might have come in different states.

The variances of visual responses were more dominated by the first dimension of activity in L2/3 than in L5 (Figure 5B-C), similar to our observations for spontaneous activity during stationary states (Figure 2A), however the slope of variance decay along the remaining dimensions was not significantly different between layers (Figure 5D-E).

Finally, we asked the extent to which the subspaces of spontaneous and sensory-evoked activity overlap in the

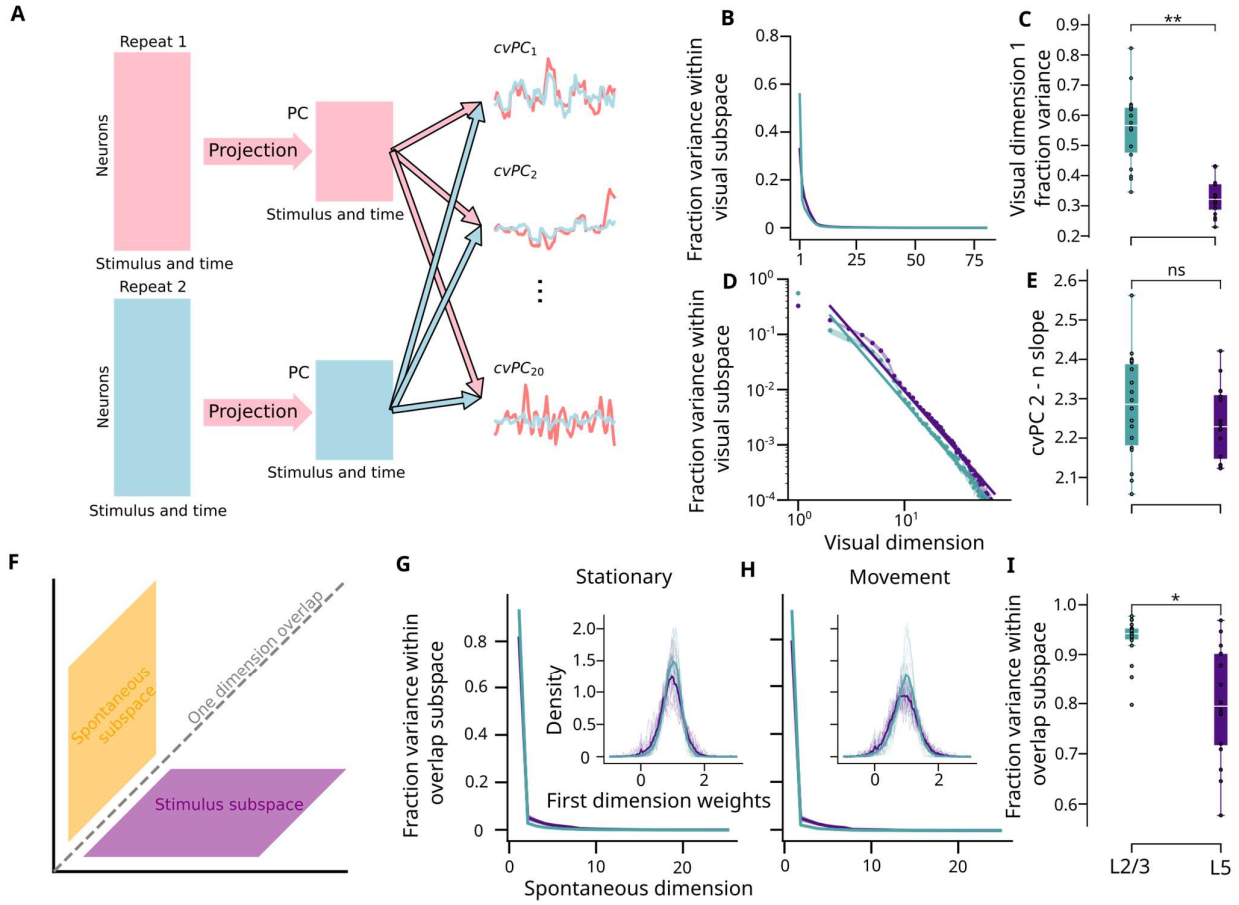

Figure 5: Population activity is more one-dimensional in L2/3 than L5. (A) Schematic of cross-validated principal components analysis (cvPCA). The matrix of stimulus-aligned responses across neurons in the first and second repeats are both projected using principal component weights learned from the first repeat. The covariance between the projected time courses of the two repeats provides an estimate of stimulus-related variance in each component. **(B)** Fraction covariance explained for each cross-validated principal component. (error bars showing standard deviation too small to be visible). **(C)** Fraction of stimulus-related variance explained by the first visual dimension in L2/3 and L5 neurons. Each dot represents a recording session. ($p = 0.007$; N = 34 sessions, linear mixed effects model). **(D)** Same as (B), but with fraction variance on a log scale, and showing the line of best fit of variance decay from the second dimension onwards. **(E)** Same as (C) but comparing the variance decay slope between L2/3 and L5 neurons. **(F)** Schematic depicting the observation of a single dimension of overlap between the movement and visual subspace in the visual cortex. **(G)** Fraction of variance of the visual subspace that is explained by each spontaneous dimension, obtained during stationary periods. Inset: distribution of weights of individual neurons on the primary shared dimension. **(H)** Same as (G), but during movement periods. **(I)** Comparison of the fraction of variance in the first overlapping dimension between spontaneous and visual activity in L2/3 and L5 neurons. * $p < 0.05$, ** $p < 0.01$

different layers. Previous work[21] showed that in L2/3, these spaces overlap almost exclusively along one dimension.

To estimate how the overlap between subspaces of spontaneous and evoked activity differed by layer, we projected visual activity onto the subspace defined by spontaneous activity, and then applied cvPCA to calculate the fraction of visual susbpace activity that can be explained by spontaneous dimensions. Consistent with previous results[21], we found that in L2/3 this overlap occurs almost exclusively in a single dimension (93% ± 1% of shared variance explained by this dimension), and the weights of neurons onto this single dimension were nearly exclusively positive (Figure 5G-I). In L5 the largest shared dimension still had nearly all positive weights, but it accounted for a smaller fraction of the shared variance (80% ± 3; Figure 5-I).

Discussion

We found multiple differences in how population coding of visual and non-visual information differs between L2/3 and L5 visual cortex. First, activity in L2/3 is more strongly modulated by visual stimuli, while activity in L5 is more strongly modulated by movement. Second, movement decreases population synchrony in L2/3, because it abolishes a spontaneous oscillation that entrains the L2/3 population, whereas movement increases population synchrony in L5, because L5 cells are less coupled to the spontaneous oscillation, and more strongly correlated with movement itself. Finally, in both layers the overlap between spontaneous and stimulus-evoked subspaces was dominated by a dimension of all-positive weights, but this dimension accounted for a larger fraction of overlap in L2/3 than L5.

Both two-photon calcium and extracellular electrophysiology are subject to biases[34]. Our results were consistent between

the two approaches with one notable exception: auditory-evoked activity was found in all layers of visual cortex in electrophysiology, but not 2p recordings. We attribute this discrepancy to the suppression of sound-evoked behavioral responses (e.g. startle) by the background noise of the two-photon microscope, which was supported by abolition of startle responses when playing recorded 2-photon noise through a speaker in a test rig. Our data thus support the hypothesis that auditory responses in visual cortex arise primarily from stimulus-evoked movements[32], and are thus stronger in L5 than L23 because L5 exhibits stronger movement correlates.

The two-photon experiments used different mouse lines to record in L2/3 and L5, raising a potential concern that observed differences between layers reflects other differences between these lines. However, our electrophysiology experiments obtained the same results in simultaneous recordings from layer 2/3 and layer 5 cells from the same mice, which mitigates this potential concern. We also found quantitative differences in the effects observed using two-photon calcium imaging and electrophysiology experiments. For example, the difference in visual responses between L2/3 and L5 was larger in 2p than electrophysiology recordings (cf Figures 5E and S5C), and population synchrony in L2/3 cells was higher when recorded with 2p than electrophysiology (cf Figure 2F and Supplemental Figure 3F). We believe these differences reflect different signal acquisition biases between imaging and electrophysiology; In particular, electrophysiology spike sorting requires relatively high baseline firing rates, decreasing its ability to record from sparsely active neurons[35] such as excitatory neurons of L2/3. The fact that visual dominance of L2/3 appeared stronger with 2p than electrophysiology might therefore reflect better 2p detection of sparsely-active neurons with low baseline rate, which are strongly dominated by vision but frequently missed by spike sorting.

All layers of mouse sensory cortex receive thalamic input[3,36,37], which carries not only visual sensory inputs but also information about behavioral state[38]. Indeed, silencing thalamus blocks the depolarization of V1 and S1 cells when mice whisk[38,39], and S1 cells receiving more thalamic input tend to be more strongly modulated by behavioral state[40]. We therefore speculate that the different balance of encoding seen in L5 vs. L2/3 neurons by movement vs. sensory stimuli we observed might reflect differences in the types of thalamic input they receive. For example, it is possible that deep-layer excitatory cells receive more inputs from higher-order thalamus – which convey cognitive signals such as movement, surprise, and locomotor mismatch[41–44] and ramify densely within L5a but not L2/3[3,45,46].

It is now clear that cortical "baseline activity" does not reflect steady firing of neurons, but rather complex patterns of coordinated firing[47–51]. Part of the structure of this cortical spontaneous activity encodes ongoing behavioral movements[21,22,32], with the remaining structure presumably reflecting cognitive processes not directly correlated with movement. When sensory stimuli are presented, cortical activity reflects an integration of sensory responses and ongoing activity, which presumably allows the cortex to interpret sensory input in the current behavioral/cognitive context. Our observation that L5 activity correlates better with movement and L2/3 with sensory stimuli suggests the balance of this sensory/cognitive integration differs between layers, with L2/3 encoding a more "pure" sensory signal and L5 a greater integration with nonsensory signals. We speculate that this different balance might relate to the primary targets of the different layers: because L2/3 outputs go mainly to higher-order cortices, coding of primarily sensory information might be suitable for areas that will continue to perform more complex sensory processing. In contrast, L5 outputs go to subcortical stations involved in motor production such as basal ganglia, superior colliculus, and brainstem, for which information that integrates sensory input with behavioral/cognitive context might be more appropriate.

References

1. Douglas, R. J. & Martin, K. A. C. Neuronal circuits of the neocortex. *Annual Review of Neuroscience* vol. 27 419–451 Preprint at https://doi.org/10.1146/annurev.neuro.27.070203.144152 (2004).
2. Thomson, A. M. & Bannister, A. P. Interlaminar connections in the neocortex. in *Cerebral Cortex* vol. 13 5–14 (Oxford University Press, 2003).
3. Harris, K. D. & Shepherd, G. M. G. The neocortical circuit: Themes and variations. *Nature Neuroscience* vol. 18 170–181 Preprint at https://doi.org/10.1038/nn.3917 (2015).
4. Harris, J. A. *et al.* Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195–202 (2019).
5. Campagnola, L. *et al.* Local connectivity and synaptic dynamics in mouse and human neocortex. *Science (1979)* **375**, (2022).
6. Petreanu, L., Mao, T., Sternson, S. M. & Svoboda, K. The subcellular organization of neocortical excitatory connections. *Nature* **457**, 1142–1145 (2009).
7. Constantinople, C. M. & Bruno, R. M. Deep cortical layers are activated directly by thalamus. *Science (1979)* **340**, 1591–1594 (2013).
8. Adesnik, H. & Naka, A. Cracking the Function of Layers in the Sensory Cortex. *Neuron* vol. 100 1028–1043 Preprint at https://doi.org/10.1016/j.neuron.2018.10.032 (2018).
9. Swadlow, H. A. Efferent neurons and suspected interneurons in S-1 vibrissa cortex of the awake rabbit: Receptive fields and axonal properties. *J Neurophysiol* **62**, 288–308 (1989).
10. Sakata, S. & Harris, K. D. Laminar Structure of Spontaneous and Sensory-Evoked Population Activity in Auditory Cortex. *Neuron* **64**, 404–418 (2009).
11. O'Connor, D. H., Peron, S. P., Huber, D. & Svoboda, K. Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron* **67**, 1048–1061 (2010).
12. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *Journal of Neuroscience* **28**, (2008).
13. Ayaz, A. *et al.* Layer-specific integration of locomotion and sensory information in mouse barrel cortex. *Nat Commun* **10**, (2019).
14. Tang, L. & Higley, M. J. Layer 5 Circuits in V1 Differentially Control Visuomotor Behavior. *Neuron* **105**, 346-354.e5 (2020).
15. Lur, G., Vinck, M. A., Tang, L., Cardin, J. A. & Higley, M. J. Projection-Specific Visual Feature Encoding by Layer 5 Cortical Subnetworks. *Cell Rep* **14**, 2538–2545 (2016).
16. Stringer, C. & Pachitariu, M. Analysis methods for large-scale neuronal recordings. *Science (1979)* **386**, (2024).
17. Vanderwolf, C. H. *An Odyssey Through the Brain, Behavior and the Mind*. *An Odyssey Through the Brain, Behavior and the Mind* (Springer US, 2003). doi:10.1007/978-1-4757-3779-0.
18. McCormick, D. A., Nestvogel, D. B. & He, B. J. Neuromodulation of Brain State and Behavior. *Annual Review of Neuroscience* vol. 43 391–415 Preprint at https://doi.org/10.1146/annurev-neuro-100219-105424 (2020).
19. Greene, A. S., Horien, C., Barson, D., Scheinost, D. & Constable, R. T. Why is everyone talking about brain state? *Trends in Neurosciences* vol. 46 508–524 Preprint at https://doi.org/10.1016/j.tins.2023.04.001 (2023).
20. McGinley, M. J. *et al.* Waking State: Rapid Variations Modulate Neural and Behavioral Responses. *Neuron* **87**, 1143–1161 (2015).
21. Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science (1979)* **364**, (2019).
22. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat Neurosci* **22**, 1677–1686 (2019).
23. Okun, M. *et al.* Diverse coupling of neurons to populations in sensory cortex. *Nature* **521**, (2015).
24. Harris, K. D. & Thiele, A. Cortical state and attention. *Nature Reviews Neuroscience* vol. 12 509–523 Preprint at https://doi.org/10.1038/nrn3084 (2011).

25. Renart, A. *et al.* The asynchronous state in cortical circuits. *Science (1979)* **327**, 587–590 (2010).

26. Altan, E., Solla, S. A., Miller, L. E. & Perreault, E. J. Estimating the dimensionality of the manifold underlying multi-electrode neural recordings. *PLoS Comput Biol* **17**, (2021).

27. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. High-dimensional geometry of population responses in visual cortex. *Nature* **571**, (2019).

28. Wekselblatt, J. B., Flister, E. D., Piscopo, D. M. & Niell, C. M. Large-scale imaging of cortical dynamics during sensory perception and behavior. *J Neurophysiol* **115**, 2852–2866 (2016).

29. Gerfen, C. R., Paletzki, R. & Heintz, N. GENSAT BAC cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron* **80**, 1368–1383 (2013).

30. Gong, S. *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *Journal of Neuroscience* vol. 27 9817–9823 Preprint at https://doi.org/10.1523/JNEUROSCI.2707-07.2007 (2007).

31. Madisen, L. *et al.* Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015).

32. Bimbard, C. *et al.* Behavioral origin of sound-evoked activity in mouse visual cortex. *Nat Neurosci* **26**, (2023).

33. Harris, K. D. *A Shift Test for Independence in Generic Time Series*. https://arxiv.org/abs/2012.06862 (2020).

34. Harris, K. D., Quiroga, R. Q., Freeman, J. & Smith, S. L. Improving data quality in neuronal population recordings. *Nature Neuroscience* vol. 19 1165–1174 Preprint at https://doi.org/10.1038/nn.4365 (2016).

35. Siegle, J. H. *et al.* Reconciling functional differences in populations of neurons recorded with two-photon imaging and electrophysiology. *Elife* **10**, (2021).

36. Sermet, B. S. *et al.* Pathway-, layer- and cell-type-specific thalamic input to mouse barrel cortex. *Elife* **8**, (2019).

37. Ji, X. Y. *et al.* Thalamocortical Innervation Pattern in Mouse Auditory and Visual Cortex: Laminar and Cell-Type Specificity. *Cerebral Cortex* **26**, 2612–2625 (2016).

38. Nestvogel, D. B. & McCormick, D. A. Visual thalamocortical mechanisms of waking state-dependent activity and alpha oscillations. *Neuron* **110**, 120-138.e4 (2022).

39. Poulet, J. F. A. & Petersen, C. C. H. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* **454**, 881–885 (2008).

40. Inácio, A. R. *et al.* Brain-wide presynaptic networks of functionally distinct cortical neurons. *Nature* **641**, 162–172 (2025).

41. Blot, A. *et al.* Visual intracortical and transthalamic pathways carry distinct information to cortical areas. *Neuron* **109**, 1996-2008.e6 (2021).

42. Furutachi, S., Franklin, A. D., Aldea, A. M., Mrsic-Flogel, T. D. & Hofer, S. B. Cooperative thalamocortical circuit mechanism for sensory prediction errors. *Nature* **633**, 398–406 (2024).

43. Neske, G. T. & Cardin, J. A. Higher-order thalamic input to cortex selectively conveys state information. *Cell Rep* **44**, 115292 (2025).

44. Petty, G. H. & Bruno, R. M. Attentional modulation of secondary somatosensory and visual thalamus of mice. *Elife* **13**, (2024).

45. Bureau, I., Von Paul, F. Saint & Svoboda, K. Interdigitated paralemniscal and lemniscal pathways in the mouse barrel cortex. *PLoS Biol* **4**, 2361–2371 (2006).

46. Koralek, K. A., Jensen, K. F. & Killackey, H. P. Evidence for two complementary patterns of thalamic input to the rat somatosensory cortex. *Brain Res* **463**, 346–351 (1988).

47. Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science (1979)* **273**, 1868–1871 (1996).

48. Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A. & Arieli, A. Spontaneously emerging cortical representations of visual attributes. *Nature* **425**, 954–956 (2003).

49. Buzsáki, G. Two-stage model of memory trace formation: A role for 'noisy' brain states. *Neuroscience* **31**, 551–570 (1989).

50. Berkes, P., Orbán, G., Lengyel, M. & Fiser, J. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science (1979)* **331**, 83–87 (2011).

51. Ringach, D. L. Spontaneous and driven cortical activity: implications for computation. *Current Opinion in Neurobiology* vol. 19 439–444 Preprint at https://doi.org/10.1016/j.conb.2009.07.005 (2009).

52. Dipoppa, M. *et al.* Vision and Locomotion Shape the Interactions between Neuron Types in Mouse Visual Cortex. *Neuron* **98**, 602-615.e8 (2018).

53. Burgess, C. P. *et al.* High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. *Cell Rep* **20**, 2513–2524 (2017).

54. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).

55. Syeda, A. *et al.* Facemap: a framework for modeling neural activity based on orofacial tracking. *Nat Neurosci* **27**, 187–195 (2024).

56. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: Flexible software for operating laser scanning microscopes. *Biomed Eng Online* **2**, (2003).

57. Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* (2016) doi:10.1101/061507.

58. Rupprecht, P. *et al.* A database and deep learning toolbox for noise-optimized, generalized spike inference from calcium imaging. *Nat Neurosci* **24**, (2021).

59. Speed, A., Del Rosario, J., Burgess, C. P. & Haider, B. Cortical State Fluctuations across Layers of V1 during Visual Spatial Perception. *Cell Rep* **26**, (2019).

Acknowledgements

We thank C. Reddy for support with surgery and lab management. M. Krumin for support with two-photon calcium imaging. R. Terry for animal breeding and maintenance. M.J. Wells for assistance in setting up the natural video protocol. A. Nunez-Elizalde for discussions during the conception of the natural video protocol. F. Takács and other members of the Cortexlab for discussions and comments. This work was supported by the Sainsbury Wellcome Centre Ph.D. program (T.P.H.S. and A.L.), Wellcome Trust (110120/185861 to P.C., 227065 to C.B., and 205093 and 204915 to M.C. and K.D.H.), ERC (101097874 to K.D.H.) and EMBO (ALTF 740-2019 fellowship to C.B.). M.C. holds the GlaxoSmithKline/Fight for Sight Chair in Visual Neuroscience.

Author contributions

T.P.H.S., P.C., K.D.H. and M.C. conceptualized the project and obtained the funds. T.P.H.S. performed the imaging experiments, C.B. and A.L. performed the electrophysiology experiments. T.P.H.S. performed the formal analysis and visualization with input from P.C. and K.D.H. T.S. wrote the original draft, and all authors contributed to the final version.

Competing interest statement

The authors declare no competing interests.

Materials and Methods

Surgery and recordings

Experimental procedures were conducted according to the UK Animals Scientific Procedures Act (1986) and under personal and project licenses released by the Home Office following appropriate ethics review.

For 2-photon calcium imaging, surgeries and recordings were performed on 10 mice, with 5 mice with the Ai94(tetO-GCaMP6s) x Camk2a-tTa x RBP4Cre genotype and 5 mice with the CamKII x tetO GCaMP6s genotype. Surgeries were performed under isoflurane anesthesia, and involved implantation of a headplate and a 3 or 4 mm circular craniotomy over the right primary visual cortex, as previously described[21,27,52]. Mice were habituated to handling and head-fixation before recordings were performed.

Natural video stimulus presentation protocol

During each calcium imaging recording, mice were head-fixed, either with their forelimbs on a steering wheel[53,54] (3/5 mice of each genotype) or on top of a treadmill (2/5 mice of each genotype). Visual stimuli were presented on three screens that cover 135 degrees in azimuth and 45 degrees in elevation, and auditory stimuli were presented using a single speaker (MF1 Multi-field magnetic speakers, Tucker-Davis Technologies). Audiovisual stimuli were presented as previously described[32], consisting of combinations of seven 4-second natural scene videos and the corresponding seven sounds, including a grey screen and no sound condition, for a total of 8x8=64 combinations. These 64 audio-video combinations were presented 4 times each in blocks, with a different order in each block, with a 2 - 4 second inter-trial-interval between each audiovisual pair chosen randomly from a uniform distribution, and a 10 minute spontaneous period between each

set of presentations. The electrophysiology recordings analyzed in this paper are the same as those analyzed in our previous study[32] . These recordings used an 11x11 combination of video and sounds, however for the analyses of the current study we subsampled only the 8 by 8 combination that matched that in imaging experiments. The electrophysiology protocol is described in full detail in our previous study[32], and we used the same data described in that study for our electrophysiological results in this study.

Movement quantification

We quantified behavioral movement by orofacial motion energy, which is better correlated with visual cortical activity than either running or pupil diameter [38]. Orofacial movements were monitored by infrared videography. An infrared LED illuminated the mouse and a camera with an infrared filter recorded at 30 frames per seconds. Videos were processed using Facemap[21,55] to estimate motion energy in a region-of-interest covering the eye, whiskers and face. The motion energy trace was down-sampled to 2.5 Hz to perform correlation analysis with 2-photon calcium imaging data. To define stationary and movement periods, we divided the instantaneous motion energy by its mean, and time bins below / above the 60th percentile of motion energy for each session were assigned to stationary / movement epochs.

Two-photon calcium imaging of layer 2/3 and layer 5 neurons

Neural activity was measured with a commercial two-photon microscope (Bergamo II, Thorlabs Inc) controlled by ScanImage[56], with a laser wavelength of 920 nm, at a scan rate of 2.5 Hz per plane across 12 planes spaced at 30 µm. Imaging was performed through either 3 or 4 mm cranial windows. After surgery, we performed widefield retinotopy mapping and used the sign field to ensure that all future 2-photon recordings were in V1. In all 2-photon experiments we performed recordings at two depth ranges: brain surface to 360 µm deep, and 180 to 540 µm below the brain surface, in alternating blocks in an A-B-A-B pattern with the depth of the first block balanced across sessions to eliminate potential biases due to imaging quality degradation or stimulus adaptation over the experiment. Both depth ranges used 12 planes at 30 µm spacing, scanned at 2.5 Hz per plane.

We obtained recordings of L2/3 or L5 excitatory populations by a combination of transgenic mouse lines and depth. L2/3 populations were recorded from CamKII x tetO GCaMP6s mice, which express GCaMP6s in all excitatory neurons [28], with only cells at depths 210 - 270 µm from the brain surface retained for analysis. Layer 5 populations were recorded from Ai94(tetO-GCaMP6s) x Camk2a-tTa x RBP4Cre mice [29,30 31], which express GCaMP6s only in layer 5 excitatory neurons (both IT and ET cells), with only cells at depths 420 - 480 µm retained for analysis; although this depth range potentially overlaps with L4, L4 cells should not be labelled in this mouse line.

The laser power P_z was shaped with depth z using the equation $P_z = P_0 e^{\left(\frac{z-z_0}{L_z}\right)}$. P_0, the percentage laser power at the most superficial depth z_0, was set in the range of 10 - 30% for superficial imaging and above 30% for deep imaging, z_0 is the depth of the most superficial imaging plane. The space constant L_z was by default set to 180 µm, but adjusted in some recordings to 300 or 500 µm to prevent saturation of recorded signals. Power was capped at 100% if the formula exceeded this value.

Extraction of pixel fluorescence signal from detected ROIs and corresponding neuropil region

To investigate and mitigate potential differences in recording quality with ROI size (which in turn can differ between layers), we investigated whether ROI size correlates with visual response reliability and noise level. We found a positive and negative correlation, respectively (Figure S6), indicating that without correction, we would find an erroneous bias towards more reliable, lower noise recordings in larger cells. To avoid this potential confound, we took the activity of each soma as the average fluorescence of the center 4 pixels of the region defined detected using Suite2p [57] (Figure S6).

Noise estimation in two-photon calcium imaging recordings

To quantify the noise level in fluorescence obtained from individual pixels or the mean fluorescence across multiple pixels corresponding to a particular region of interest, we first computed the $\Delta F/F$ of the fluorescence signal:

$$\Delta F/F = \frac{F - F_0}{F_0}$$

where F is the fluorescence signal and F_0 is the mean fluorescence across time for the entire recording. We then computed the noise metric ν introduced in Ref. [58]:

$$\frac{\sigma_{\Delta F/F}}{\sqrt{f_r}} = \frac{\text{Median}_t |\Delta F/F_{t+1} - \Delta F/F_t|}{\sqrt{f_r}}$$

where ν is the estimated noise level, $\Delta F/F_{t+1} - \Delta F/F_t$ is the difference in signal between subsequent frames, and f_r is the frame rate of the recording (which is 2.5 Hz in all recordings in the present analysis). All analyses using based on comparing properties of single cells (Figures 1, 4, S1, S3, S4G) only analyzed neurons with a noise level of 25% or below. Analyses of population geometry (Figures 2, 3,5) did not apply the threshold, to ensure larger population size.

Estimating layer depth from Neuropixels recordings

We estimated laminar depth in Neuropixels recordings following Ref. [59], using the code at https://github.com/haiderlab/Current-Source-Density. Briefly, current source density analysis (CSD) was performed on the average local field potential aligned to video onset. The top of layer 4 was identified as the earliest and largest sink in CSD, and the depth of all units is defined relative to the top of layer 4.

Separating visual and auditory-evoked components

To isolate the visual-related and auditory-related components of each cell's fluorescence response we used an ANOVA-like decomposition [32]. Each neuron's response can be written as:

$$f(t, a, v) = g(a, t) + h(v, t) + q(a, v, t) + \beta(t) + \epsilon$$

where $f(t, a, v)$ is a cell's signal as a function of time t, auditory stimulus a and visual stimulus v; $g(a, t)$ is the auditory component of the cell's response as a function of auditory stimulus and time; $h(v, t)$ is the visual component of the cell's response as a function of the visual stimulus and time; $q(av, t)$ is the interaction component of the cell's response as a function of the audiovisual stimuli and time, $\beta(t)$ is the mean component of the cell's response independent of stimulus identity, and ϵ is the trial-to-trial variability that is not modelled.

Population coding in cortical layers

To estimate the visual and auditory components of stimulus responses, we first z-scored each cell's activity relative to the ensemble of pre-stimulus activity (all timepoints within 0-1 s prior of any trial) to yield an array $f(t,a,v)$ for each cell, whose pre-stimulus activity had zero mean by definition. We then estimated the grand mean component of each cell's response by taking the average activity over all videos and sounds:

$$\hat{\beta}(t) = \frac{1}{N_v N_a} \sum_v \sum_a f(t,a,v)$$

Where $\hat{\beta}_t$ is the estimated mean response, N_a is the number of auditory stimuli, and N_v is the number of visual stimuli.

To obtain the visual-related component, we take the cell's response over all possible sounds, and subtract that by the mean response component:

$$\hat{h}(v,t) = \frac{1}{N} \sum_a f(t,a,v) - \hat{\beta}(t)$$

To obtain a single number quantifying how much each neuron was driven by visual stimuli (Figure 4), z-scored $\hat{h}(v,t)$ to an ensmble of that neuron's activity in the range t=-1 to 0s prior to stimulus onset. We then take the average response timecourses across all videos, to yield a video-evoked response timecourse for each unit over the range 0 – 4 seconds relative to stimulus onset (Figure 4C), and summarized this by avaraging over time (Figure 4D).

We used the same approach to obtain the auditory-related component (Figure S4).

Linear shift test of movement correlation

To test for a significant correlation between the time courses of movement and neural activity (Figures S1, S3), we performed a linear shift test[33]. For each neuron, we computed the Pearson correlation coefficient r between its time course and the movement activity time course with shifts from -200 to 200 frames (1.2 seconds per frame). We then use a p-value threshold of 0.05 to indicate statistical significance, detecting a significant correlation if the unshifted r value is below the 2.5 or above the 97.5 percentile of the ensemble of all r values.

Shared variance components analysis

The quantification of overlap between spontaneous and visual subspaces was performed as in Ref. [21] . To obtain estimates of the spontaneous variances we applied shared variance component analysis (SVCA) separately during stationary and movement periods (Figure 3). The fluorescence signal of each cell was first z-scored across the entire recording. We then partitioned the cells into a train and test set (with number of cells $N_1$and N_2) by dividing the imaging field-of-view into a 4 by 4 grid, and taking cells from alternating grid squares. We also partitioned time points into train and test set (with number of time points T_1 and T_2) by taking alternating 60 second segments. This yielded 4 activity matrices: $S_1 \in \mathbb{R}^{N_1 \times T_1}, S_2 \in \mathbb{R}^{N_2 \times T_1}$containing the activity of cell set 1 and 2 during the train set time points, and $S'_1 \in \mathbb{R}^{N_1 \times T_2}, S'_2 \in \mathbb{R}^{N_2 \times T_2}$ containing the activity during the test set time points. We then compute the covariance between cell set 1 and cell set 2 during the train time points and performed singular value decomposition with rank r on this covariance matrix to obtain our shared variance components (SVC) for cell set 1 in $U \in \mathbb{R}^{N_1 \times r}$ and cell set 2 in $V \in \mathbb{R}^{N_2 \times r}$

$$S_1 S_2^\intercal = USV^\intercal$$

We then projected the corresponding cell set's activity during the train time points on these shared variance components

$$S'_{1,\text{proj}} \in \mathbb{R}^{r \times T_2} = U^\intercal S'_1,$$
$$S'_{2,\text{proj}} \in \mathbb{R}^{r \times T_2} = V^\intercal S'_2$$

The shared variance along each SVC was estimated by the covariance between these two sets of projected cell activity during the train time points: $S'_{1,\text{proj}} S'^{\intercal}_{2,\text{proj}}$

Cross-validated principal components analysis (cvPCA)

To estimate the variance in successive dimensions of signal coding subspace we used cross-validated principal components analysis (cvPCA) [27].

We divided the repeats of each stimulus into two halves, and used each half i to obtain a matrix $X^{(i)} \in \mathbb{R}^{N_t N_v \times N_c}$ containing the concatenated timecourses of all stimulus responses, z-scored relative to pre-stimulus baseline activity as described above, for all cells (N_t is the number of time points aligned to stimulus onset, N_v the number of videos, and N_c the number of cells). We performed principal components analysis on the first activity matrix $X^{(1)} \in \mathbb{R}^{N_t N_v \times N_c}$ to obtain the principal components of visual activity, and projected the second-half of the data $X^{(2)}$ onto these components, and calculated the covariance between the corresponding projections to obtain estimates of the visually-driven variances in the population (Figure 5B-E).

Spontaneous and visual shared subspace analysis

To estimate the overlap between the spontaneous and visual subspaces (Figure 5G-I), we used the method of Ref. [21]. Visual activity $X \in \mathbb{R}^{N_t N_v \times N_c}$ was projected onto the first 25 components of the spontaneous subspace identified with SVCA: $M \in \mathbb{R}^{N_c \times k} = [UV]_{:,1:k}$, where $k = 25$, and cvPCA was performed on these projected components. The variance explained by each component quantifies the overlap between visual and spontaneous subspace. The weights of each neuron on the first shared dimension were obtained by calculating the linear combination of weights of the first shared dimension with the weights of each neuron on each of the first 25 spontaneous dimensions:

$$w_1 \in \mathbb{R}^{N_c} = MP_{:,1}$$

Where $P \in \mathbb{R}^{k \times p}$ is the loading of each the spontaneous subspace onto the principal components p obtained via cvPCA.

The weights were scaled for display to have a mean-squared value of 1 (Figure 5G-H, insets).

Population coupling index

To quantify the synchrony of population activity, we computed the population coupling index of each neuron, defined as the Pearson's correlation of the neuron's activity with the mean population response:

$$\text{PCI}_\text{i} = \frac{\text{Cov}(\text{x}_\text{i}(\text{t}),\ \mu(\text{t}))}{\sigma_{\text{x}_\text{i}(\text{t})} \sigma_{\mu\,(\text{t})}}$$

Where $x_i(t)$ is the activity of neuron i across time, and $\mu(t)$ is the mean activity of all neurons across time.

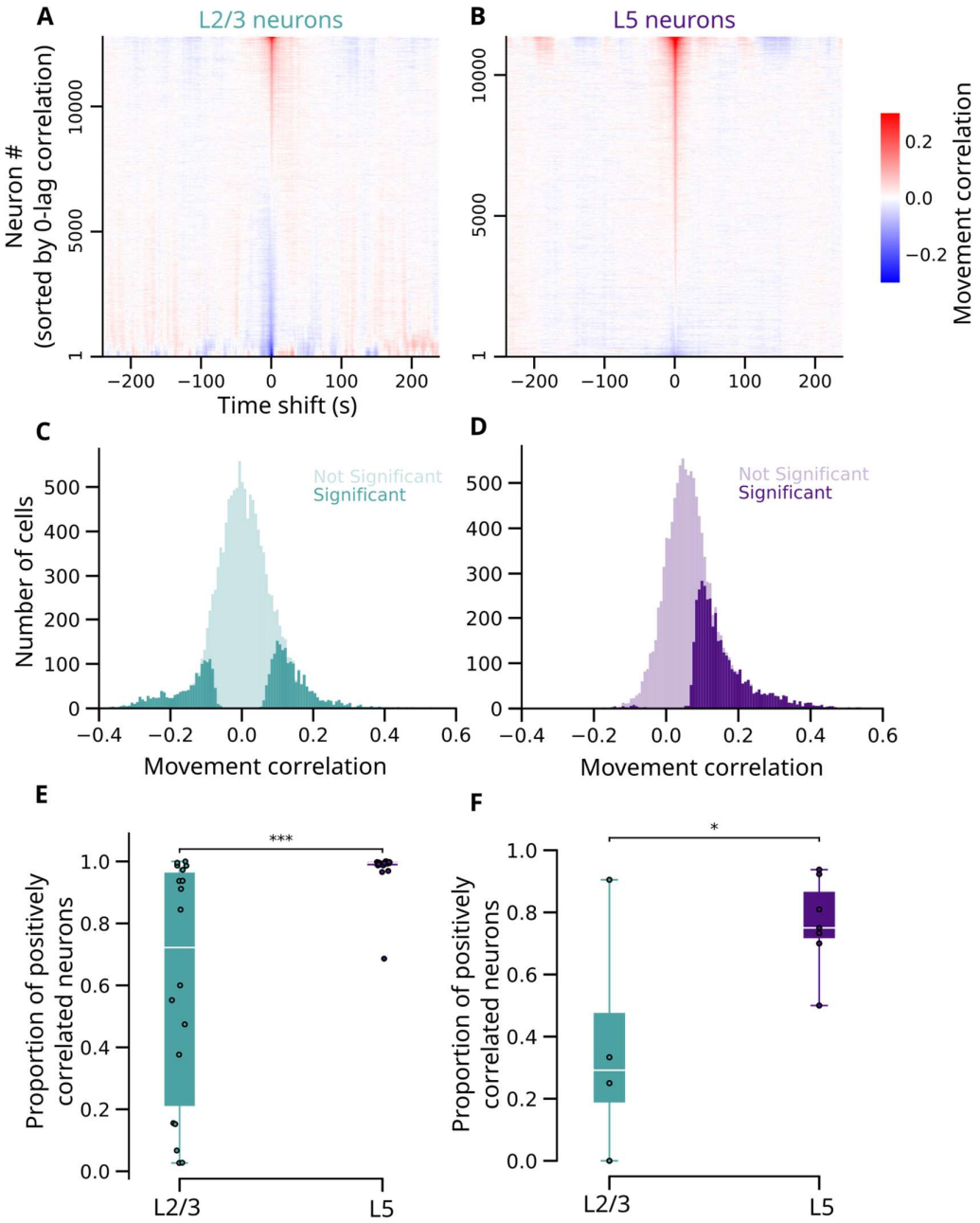

Figure S1: Testing neurons for significant movement correlation with the linear-shift test. (A) Heatmap of Pearson correlation values between fluorescence from neurons in L2/3 and lagged motion energy time courses, with different time shifts. Neurons are combined from all sessions and arranged vertically sorted by lag-0 correlation coefficient. **(B)** Same as (A) but for L5 neurons . **(C)** Histogram of correlation values in all recorded L2/3 neurons with significant (dark) and insignificant (light) correlation as assessed by the linear shift test. **(D)** Same as (C) but for layer 5 somas. **(E)** Proportions of significantly correlated neurons whose correlation is positive, in L2/3 and L5. Dots: individual recording sessions. The fraction in L2/3 and L5 are significantly different ($p < 0.001$, linear mixed effects model). **(F)** Same as (E), but for Neuropixels recordings ($p < 0.05$, linear mixed effects model).

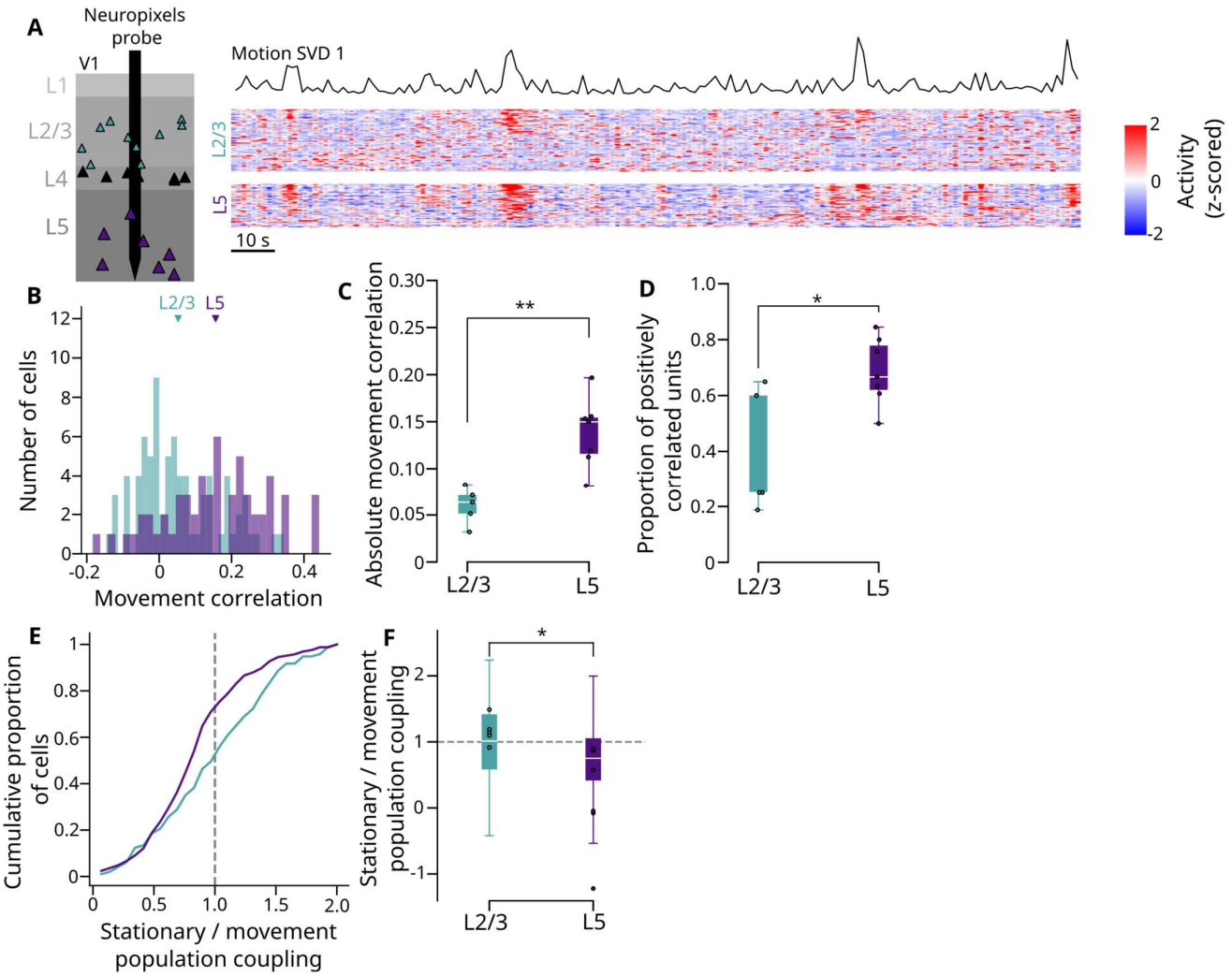

Figure S2: Movement correlation and population coupling in electrophysiology recordings. (A) Example recording using Neuropixels probes simultaneously in L2/3 and L5. Within each layer group, cells are sorted by signed movement correlation. **(B)** Histogram of signed movement correlation values of neurons in (A), including significant and non-significant correlations. **(C)** Comparison of the mean absolute movement correlation (significant or not) of neurons during each recording session in L2/3 and L5 (Difference between layers: $p = 0.01$, Wilcoxon rank-sum test, L2/3: N = 5 sessions; L5: N = 7 sessions). **(D)** Comparison of the proportion of cells positively correlated with movement (significantly or not) during each recording session (Difference between layers: $p = 0.03$, Wilcoxon rank-sum test, L2/3: N = 5 sessions; L5: N = 7 sessions). **(E)** Cumulative histogram of the ratio between the population-coupling index during stationary and movement periods across all recorded cells using Neuropixels. **(F)** Comparison of the population coupling index ratio in L2/3 and L5 ($p < 0.01$, linear mixed effects model). * $p < 0.05$, ** $p < 0.01$

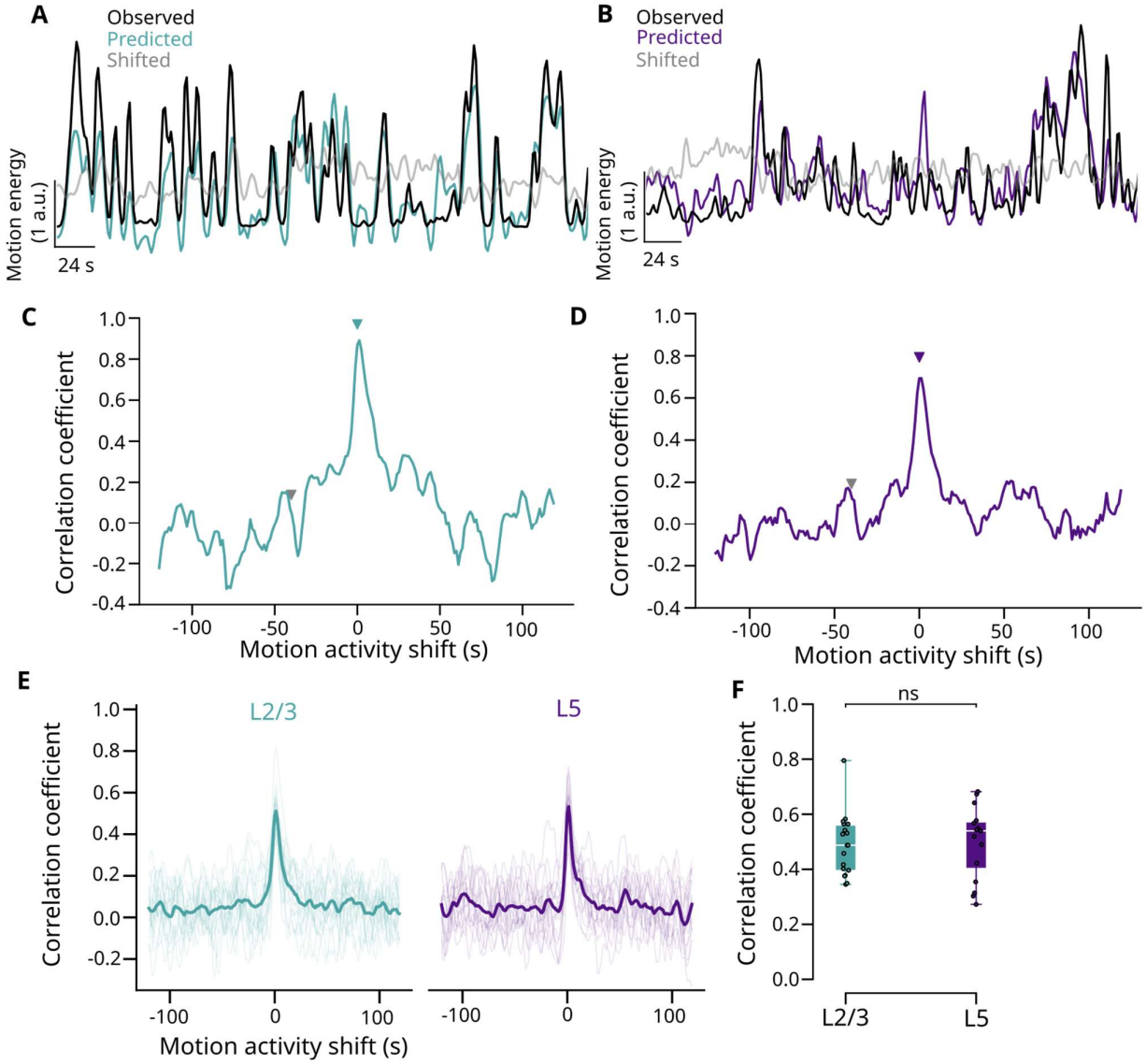

Figure S3: Regression of motion energy using neural activity. (A) Example motion energy trace (black) and motion energy predicted from L2/3 neural activity (green), for an example 2-photon recording. Grey line: prediction from neural activity time-shifted by -40 seconds (so neural activity lags). **(B)** Same as (A) but for an example 2-photon recording from L5. **(C,D)** Predictability of motion energy from neural activity, quantified by the Pearson correlation between observed and predicted motion energy, as a function of time shift, for these two example sessions. **(E)** As in (C-D), for all sessions (faint lines). Solid line: mean over sessions. **(F)** Comparison of zero time lag predictability for L2/3 and L5 (no significant difference between layers; $p > 0.05$), rank sum test across sessions.

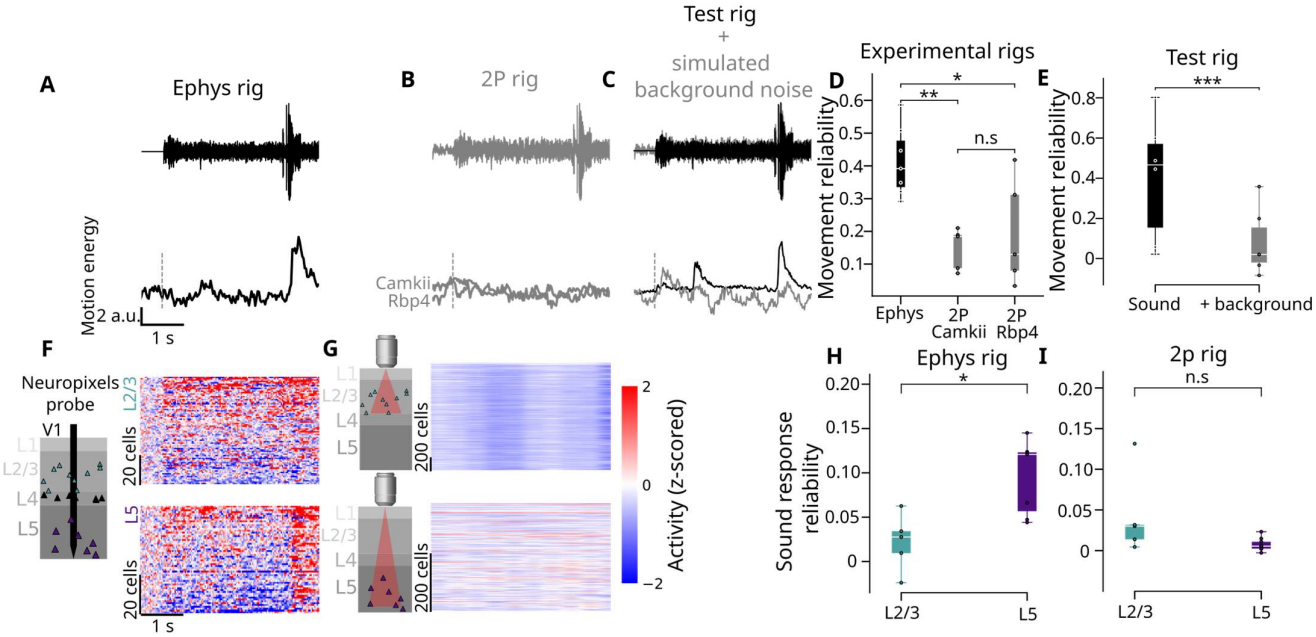

Figure S4: Differences in sound-evoked movement and neural response between electrophysiology and imaging recordings (A) Top: envelope of an example sound played during electrophysiology recording. Bottom: corresponding motion energy trace averaged across trials from one example mouse. **(B)** Top: sound envelope including background noise in the two-photon imaging rig. Bottom: corresponding motion energy traces averaged over trials, from two example mice (one Camkii and one Rbp4). **(C)**. Top: sound envelopes representing the sounds played in the test rig, with and without simulated background noise. Bottom: trial-average motion energy traces with (gray) and without (black) background sound in one example session. **(D)** Boxplots comparing the reliability, quantified as the mean pairwise correlation of the movement traces across repeats, of auditory-evoked movements in mice recorded in the electrophysiology and imaging rigs. Sound-evoked movements were significantly more reliable in electrophysiology than 2-photon recordings ($p = 0.004$ between Ephys and 2P-Camkii, $p = 0.042$ between Ephys and 2P-Rbp4, $p = 0.917$ between 2P-Camkii and 2P-Rbp4, rank sums test across mice, N = 7 ephys mice, N = 5 for both Camkii and Rbp4 mice). **(E)** Boxplot comparing the reliability of auditory-evoked movements in the test rig with and without background sounds ($p < 0.001$, linear mixed effects model, N= 6 mice). **(F)** Pseudocolor raster for example L2/3 (top) and L5 (bottom) Neuropixels-recorded population, showing neural responses to the same sound as in (A) with aligned time axes. **(G)** Same as (F) but for example 2-photon session in L2/3 (top) and L5 (bottom), aligned to trace in (B). **(H)** Boxplot comparing the reliability of sound-evoked neural responses in layer 2/3 and layer 5 neurons acquired through electrophysiology ($p < 0.05$, rank sum test, N = 5 sessions in L2/3 and N=7 sessions in L5) **(I)** Same as (H) but for imaging recordings. ($p > 0.05$, rank sum test, N = 5 for both Camkii and Rbp4 mice) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

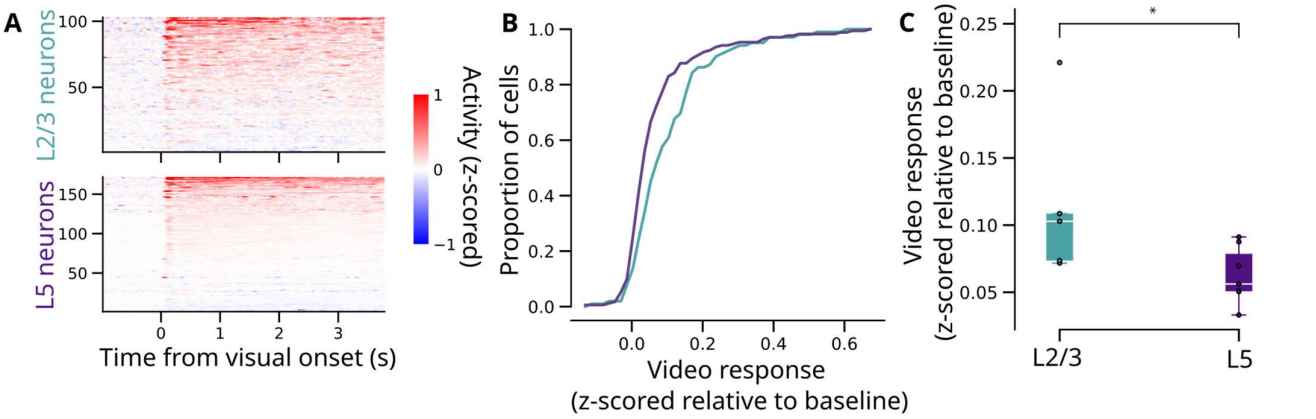

Figure S5: Visual responses in electrophysiology recordings. (A) Example recording using Neuropixels probes simultaneously in layer 2/3 and layer 5. Within each layer group, the cells are sorted vertically by their signed visual response. **(B)** Cumulative histogram of mean visual response of neurons in the example sessions of (A). **(C)** Comparison of the mean visual response of neurons in L2/3 and L5. Each point represents the mean visual response (mean across all sounds and from 0 – 4 s aligned to video onset) of all cells in a recording session. Visual responses are significantly larger in L2/3 than L5 ($p < 0.05$, linear mixed effects model, N = 5 sessions in L2/3 and N = 7 sessions in L5). * $p < 0.05$

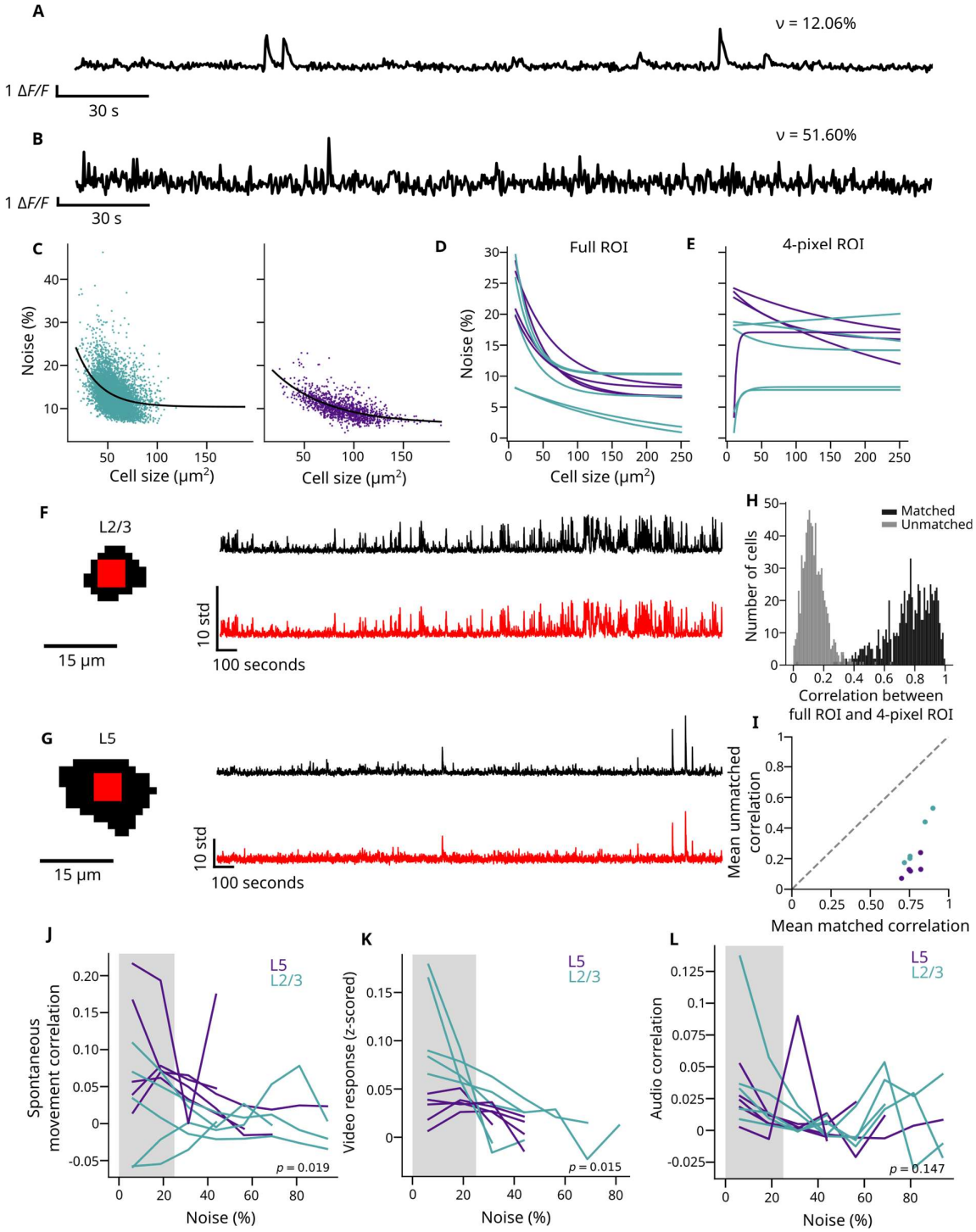

Figure S6: Controlling for the effect of soma size on noise levels. (A-B). Activity traces of two example L2/3 neurons with low and high estimated noise levels, together with estimates of noise metric ν. **(C)** Relationship between cell size and noise level of traces extracted from full ROIs for example sessions in L2/3 (left) and L5 (right). Black curve: exponential curves ($y = Ae^{-x/b} + C$) fit to these points by the Levenberg-Marquardt algorithm. **(D)** Superimposed curves showing exponential fits of relationship between cell size and noise level of traces extracted from full ROIs, for all sessions in L2/3 (green) and L5 (purple), each line represents the exponential decay curve fit to neurons obtained from each mouse. **(E)** Same as (D), but after controlling for size by using equally-sized ROIs consisting of the central 4 pixels from each ROI defined by suite2p. **(F)** Left: Silhouette of one example cell in L2/3 extracted using suite2p (black) and the center 4 pixels (red). Right: corresponding activity traces. **(G)** Same as (F) for an example neuron in L5. **(H)** Histogram over all ROIs of correlation coefficients between the full ROI activity and the 4-pixel ROIs activity (matched, black). Gray: histogram of full ROIs and 4-pixel ROIs of other units (cell identity shifted by 1; unmatched, gray). **(I)** Mean correlation between matched and unmatched full and 4-pixel ROIs for each mouse recorded in L2/3 (green) and L5 (purple). **(J)** Mean spontaneous movement correlation of each 4-pixel ROI in L2/3 (green) and L5 (purple) as a function of noise level. The gray region depicts the noise level range used in the cell-level comparisons in this study (Figure 1, 4, S4I). **(K)** Same as (J) but for video response (z-scored relative to baseline). **(L)** Same as (J) but for auditory response correlation.