Population responses of mouse visual areas to dynamic stimuli

Sahil Arora, Sean O'Connell, Andrew Sedler, and Erin Shappell

Abstract

Until recently, activity in the motor cortex has been thought to represent external movement parameters. The emerging field of neural population dynamics has since provided a comprehensive framework for explaining and studying the temporal evolution of neural activity in the motor cortex, especially during prepared reaches [1]. While the motor cortex is responsible for generating patterns that result in movement, activity in visual areas is primarily driven by inputs, so substantial efforts have been made to understand the responses of individual neurons to characteristics of stimuli. Recent work has shown that ensembles with recurrent excitatory connections, rather than single neurons, are the functional units of activity in the visual cortex and that they are activated in response to stimuli [2]. The importance of neural ensembles has also been supported by evidence of their long-term stability in the mouse visual cortex [3]. In this work, we incorporate the concepts of population dynamics and neuronal ensembles to investigate the population responses of mouse visual areas to drifting gratings and natural movies. We find low-dimensional activity in the mouse visual cortex in response to drifting gratings that contains information about stimulus conditions and is consistent across animals. We also find that population states in subsets of V1 are predictive of the spiking activity of neurons in other subsets. For natural movies, we find similar predictive relationships between areas and also find that we can align activity in visual areas across animals, suggesting that the recorded populations share a common structure even for complex stimuli. These results provide evidence that the dynamics of overlapping neuronal ensembles, rather than individual neurons, are the building blocks of visual processing.

Introduction

Modern neuroscience research is beginning to point heavily to the importance of studying coordinated activity of large neural populations rather than investigating individual neuron firing patterns in the control of behavior [4]. This new perspective is not only being applied in motor control literature, however. For example, it has recently contributed greatly to the understanding of the hierarchy of visual processing, such as in viewpoint invariance during object recognition [5]. It can also offer insight into understanding early visual processing of simpler features like contrast, edge detection, and scene movement.

At the core of many of these techniques is *dimensionality reduction*, where the firing rates of hundreds of neurons can be condensed to a smaller number of features, referred to as *latent features*. This operation can be thought of as a transformation from some high-dimensional

space (spanned by the entire population) down to a lower-dimensional space, or *latent space*. Depending on the methods used (e.g. factor analysis (FA), principal component analysis (PCA), demixed PCA (dPCA), canonical correlation analysis (CCA), etc.), the latent features will reflect a compressed form of the information embedded in the higher-dimensional activity. Each method will extract a different aspect of the information in the population activity, so the method should be chosen according to the information desired by the researcher. For example, PCA is a common dimensionality reduction method for projecting neural population data onto axes aligned to state space directions that capture the most variance in the dataset, which is usually related to the amount of information along that dimension. However, a potential drawback is that PCA treats all variance in the dataset the same, including noise or trial to trial variability that is shared across the population. FA is ideal for dealing with these problems [6]. It extracts underlying features in the neural population data while also modeling the contributions of noise in each recording channel, which allows isolation of more representative latent factors. dPCA can be used to extract task-specific low-dimensional factors, and CCA operates to find maximal shared structure across separate observations of similar high-dimensional data.

Throughout this study, our primary aim was to investigate the nature of neural population responses to different types of dynamic visual stimuli (various drifting gratings and natural movies). We apply several of these recent dimensionality reduction techniques to investigate how well we can extract specific aspects of the visual information from the higher dimensional neural space across different visual cues and even different mice. Because of the complexity of natural visual stimuli, our initial experiments focus on neural responses to several parametrized simple visual stimuli: drifting gratings at different angles, movement speeds, and spatial frequencies. After this line of analysis, we investigate the latent structures of the neural responses to the more complex natural movies, both within and across different mice.

All of these goals can be ideally approached by applying the most well-suited dimensionality reduction technique. For example, we use FA for many of our analyses in this study to explore how well the latent factors in one brain area linearly correlate with the spiking activity in other areas. In a separate analysis, we apply dPCA to extract meaningful components of the high-dimensional neural activity that correspond with parameters of the external visual stimulus. Our desire in using dPCA was to observe how well the information encoding different features of the visual stimulus could be extracted, and how this information was distributed across brain areas. Our final goal was to investigate the strength of shared structure in the latent spaces across different mice exposed to the same natural movies. We used CCA to expose the optimal shared latent structures at several different levels of dimensionality reduction (varying dimensions of the latent space). Across our range of experiments and dimensionality reduction approaches, the significance and explanatory power of these neural latent spaces for understanding visual processing within the mouse visual cortex (V1) is evident.

Methods

Data Collection and Preprocessing

Data was collected as part of a large-scale initiative by the Allen Institute for Brain Science and released in 2019 as part of the Allen Brain Observatory project [7]. Here, we give a brief overview of the relevant data collection protocols, but refer the reader to the technical whitepaper for more details [8]. Wild-type and transgenic C57BL6/J mice were surgically implanted with multiple Neuropixel probes in visual areas identified by intrinsic signal imaging. After habituation, the animals were mounted on an experimental rig and passively viewed sets of stimuli, including 2-second exposures to drifting gratings for all combinations of 8 orientations and 5 temporal frequencies (drifting gratings, 14-15 trials / condition / session) and two short clips of footage from the opening scene of the movie *Touch of Evil* (natural movies, 10 or 20 trials / condition / session). Simultaneous neural activity from a variety of visual brain areas was recorded via Neuropixel probes for the duration of the stimuli.

VISUAL CORTEX	primary visual cortex	VISp	3964[1] (8603[2])
	lateromedial area	VISl	2075 (4935)
	rostrolateral area	VISrl	2567 (6013)
	anterolateral area	VISal	3036 (6466)
	posteromedial area	VISpm	1798 (4215)
	anteromedial area	VISam	2959 (6198)
HIPPO-CAMPAL FORMATION	cornu ammonis 1	CA1	5878 (17,104)
	cornu ammonis 3	CA3	815 (3148)
	dentate gyrus	DG	1655 (5832)
	subiculum	SUB	850 (1938)
	prosubiculum	ProS	652 (1522)
THALAMUS	lateral geniculate nuc.	LGd	1306 (2582)
	lateral posterior nuc.	LP	2492 (4849)
MIDBRAIN	anterior pretectal nuc.	APN	1297 (3841)

[1]Total units passing default QC filters
[2]Total units (no QC filters)

Table 1. The total number of recorded units from each major brain structure recorded in the dataset, along with groupings of the areas into larger functional groups.

We chose 12 brain structures with largest numbers of recorded units (see Table 1, excluding subiculum and prosubiculum) and selected five sessions recorded from different animals that included data from all of these areas. For some analyses, we grouped neurons from the 12 structures into four distinct functional areas, also shown in Table 1 - visual cortex, hippocampus, thalamus, and midbrain. For collaborative efficiency, we stored the data files and performed all analyses using a shared virtual machine in Google Cloud. For all analyses, we binned single-trial spiking activity aligned to stimulus presentation at 1 ms (including margins to eliminate edge effects) then smoothed the data by convolving with a 60 ms Gaussian kernel. Finally, we re-binned the data to 10 ms and used the resulting single-trial rate estimates in all subsequent analyses.

Extracting Condition-related Activity

To extract condition-independent and condition-dependent components of the neural activity during the drifting grating experiments, we first computed peristimulus time histograms (PSTHs) for all 40 conditions by averaging the single-trial rate estimates for each neuron. We extracted only neurons from the visual cortex (100-200 neurons) and applied demixed principal components analysis (dPCA) to extract the components that best separated the activity with respect to time, grating angle, and temporal frequency. We performed this analysis across all 5 sessions and achieved similar results, but we only show one session here. We also examined the weights of the dPCA loading matrix for an indication of which areas of the visual cortex were important for distinguishing between conditions.

Latent State Alignment Across Sessions

To align latent activity across sessions, we used a method described in an existing paper for aligning motor cortical dynamics across sessions on different days [9]. We first computed visual cortex PSTHs for all 40 drifting grating conditions for each session as described above. We then extracted 10-dimensional unaligned latent dynamics by separately applying PCA to each session. Finally, we used canonical correlations analysis (CCA) to find the linear projections of the latent dynamics that were maximally correlated, which we call aligned dynamics. We computed the absolute Pearson's *r* for unaligned and aligned dynamics and averaged across dimensions and repeated for all pairs of the 5 sessions. We also performed the analysis in 3D on one pair of sessions for visualization purposes.

Spike Prediction

To examine the relationships between population activity in the 12 brain structures, we separately looked at neural activity during one drifting grating condition (180°, 2 Hz) and one natural movie condition (30s). We extracted 10-dimensional latent states from the single trial rate estimates for each structure using Factor Analysis (FA). We then trained Poisson regressors to predict the Poisson-distributed spiking activity of all observed neurons based on the latent state of a given structure. We evaluated the performances of these models using pseudo-R^2, taking the structure-to-structure performance as the median pseudo-R^2 across all neurons in the target structure. For a baseline comparison, the approach was replicated with ridge regression in place of the Poisson regressors. This is a linear least squares method that applies an L2-regularization penalty to the data. Instead of fitting a Poisson distribution on the latent factors, a regularized line was fitted on the smoothed PSTHs to predict spiking activity across structures and neurons.

Population Alignment Across Sessions

To perform canonical correlation analysis (CCA) on the Allen Institute data, we trained and tested CCA models using SciKit Learn's CCA method [10]. CCA has been used to capture an underlying pattern within two sets of data and align the data in a smaller latent dimensionality.

We used CCA to determine how accurately two sets of neural responses from different mice and/or to different stimuli could be aligned, or how strong of an underlying pattern exists within the two sets. We used the coefficient of determination (R^2) to evaluate each alignment's accuracy.

Because of the promising results from aligning neural responses to drifting gratings using CCA, we chose to continue to investigate common patterns within visual response data by moving to a more complex stimulus: natural movies. We examined neural responses to the Allen Institute's two natural movies (labeled "natural movie one" and "natural movie three"), which are clips taken from the movie "Touch of Evil." Because the two natural movies were different lengths (natural movie one was 30 seconds and natural movie three was 120 seconds), we only aligned neural responses to the first 30 seconds of natural movie three. We ran a total of three experiments: the first two aligned neural data across all 12 brain structures of interest, and the third aligned neural data for only the primary visual area (VISp). In the first and third experiments, we chose a lower latent dimensionality of 3, and in the second experiment we chose a higher latent dimensionality of 20 to determine if the larger input data size from examining across all areas affected the alignment accuracy.

For each of the three experiments where we examined neural responses across all brain areas or on one brain area, we completed three sub-experiments. In the first sub-experiment we aligned neural responses to natural movie one from five different sessions. In the second sub-experiment we aligned neural responses to natural movie three from the five sessions. In the third sub-experiment we attempted to align neural responses to natural movie one to responses to natural movie three from the five sessions to serve as a lower R^2 reference, since we expected the alignment to be significantly poorer when changing two variables. In sub-experiments 1 and 2, we included cases where the response data was aligned with itself as an upper R^2 reference since we expected the alignment to be perfect ($R^2 = 1.0$). For each sub-experiment, a total of 25 CCA models were trained and evaluated (one model for each alignment pairing across the five sessions).

Results and Discussion

Drifting Gratings

V1 shows more interpretable population activity than other visual areas

Of all functional areas examined, the PSTHs from the visual cortex display the most structured response to the various drifting grating conditions. The population activity starts at a similar region in state space for all conditions. In the 200 ms after the stimulus is presented, the state rapidly jumps in a similar direction for all conditions, then settles into a stable state unique to each condition for the remainder of the stimulus presentation. Some interesting findings from this simple analysis are the similarities in representations of gratings moving in opposite directions (e.g. the gratings moving left and right have very similar trajectories and stable

points). Additionally, the trajectories for up-down gratings are much more distinct from the other pairs, indicating that the visual cortex represents up-down motion much differently from any motion with a horizontal component. It makes sense that a mouse would represent this information differently from an evolutionary standpoint, and this is consistent with the findings of others [11].

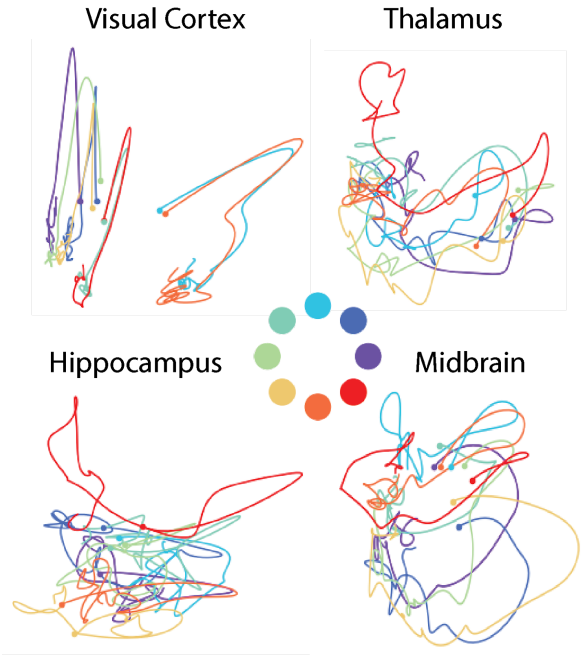

Figure 1. Latent population activity extracted by FA for all eight drifting grating orientations at a temporal frequency of 4 Hz. The legend in the middle shows the colors for various drift directions.

On the other hand, the other functional areas do not appear to show much consistent structure related to the stimulus condition. This is likely because these areas have functions that are less directly related to the observed visual stimulus at any given point in time, with the thalamus regulating attention, hippocampal areas managing memory formation, and the midbrain managing autonomic responses to ambient light.

V1 contains information about stimulus condition

When we examine the neural data for variability that separates the various condition types for the drifting grating stimulus using dPCA, we find that representations are cleanly separated for all 40 conditions across all sessions, with the exception of overlapping representations for opposite grating directions. Also, we again find that states corresponding to the up-down stimuli are very distinct from all other directions.

If V1 were strictly representational, one might expect to see oscillatory activity that corresponds to the oscillating visual field. However, viewing these results from a dynamical systems perspective, V1 seems to demonstrate attractor-like dynamics for each condition, where the precise location of the attractor in state space is mediated by the contextual input of a visual stimulus.

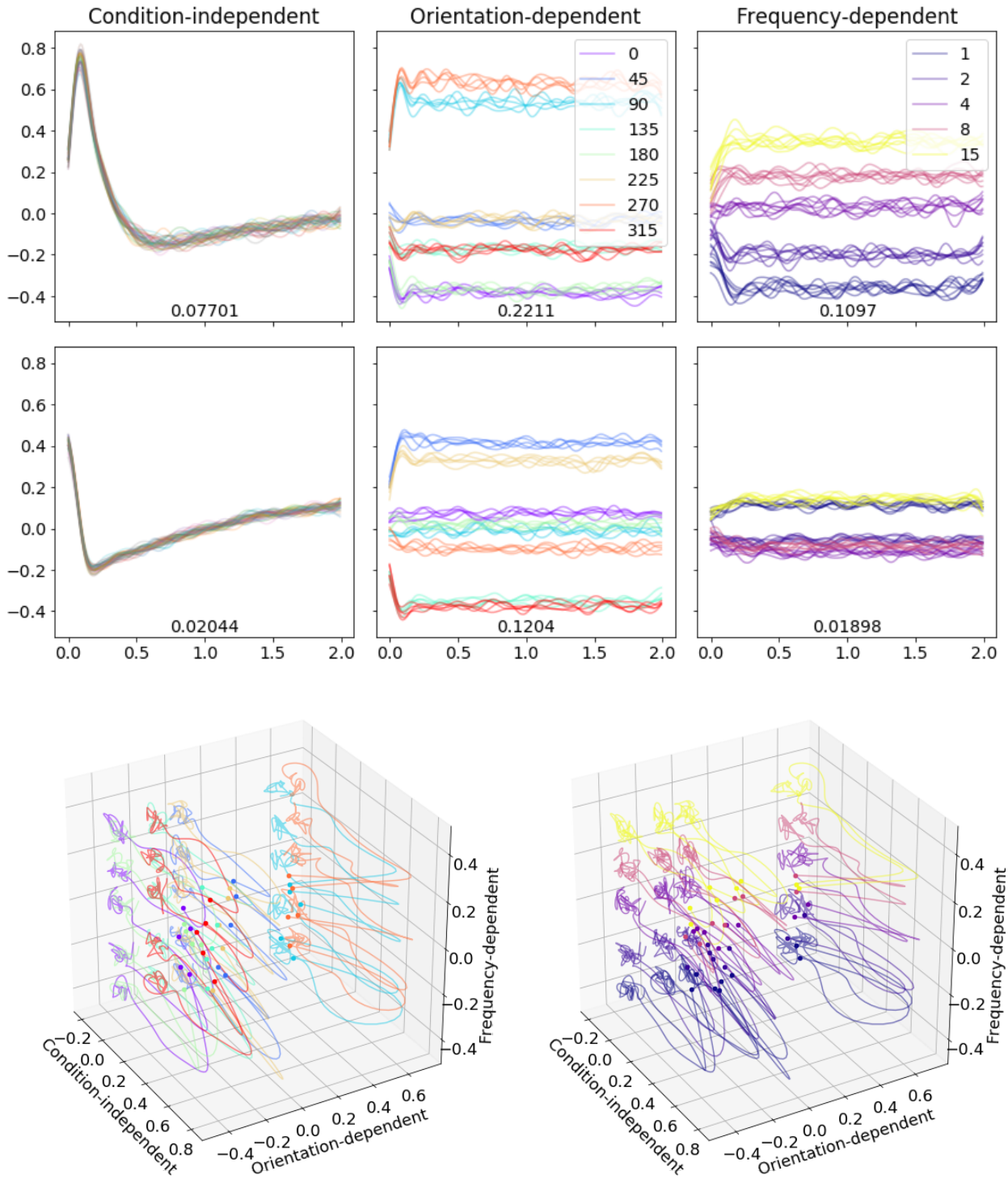

Figure 2. Results of applying dPCA to separate visual cortex PSTHs from 40 behavioral conditions. (a) Single-dimension plots of the first two components of time-varying activity, orientation components, and temporal frequency components. Numbers indicate percent of variance explained by each dimension. (b,c) Three-dimensional visualizations, colored by stimulus orientation and by temporal frequency.

When we performed this analysis across other functional areas, we saw similar time-varying components in the thalamus and midbrain (7-8% of variance) but not in the hippocampus, virtually no separation of drift angle in any of these regions, and mild separation of temporal frequency components in all areas (9-12% of variance). This is consistent with our understanding of the response of the thalamus and midbrain to a change in stimulus, and the potential sensitivity of all areas to higher temporal frequencies which may trigger a fear response.

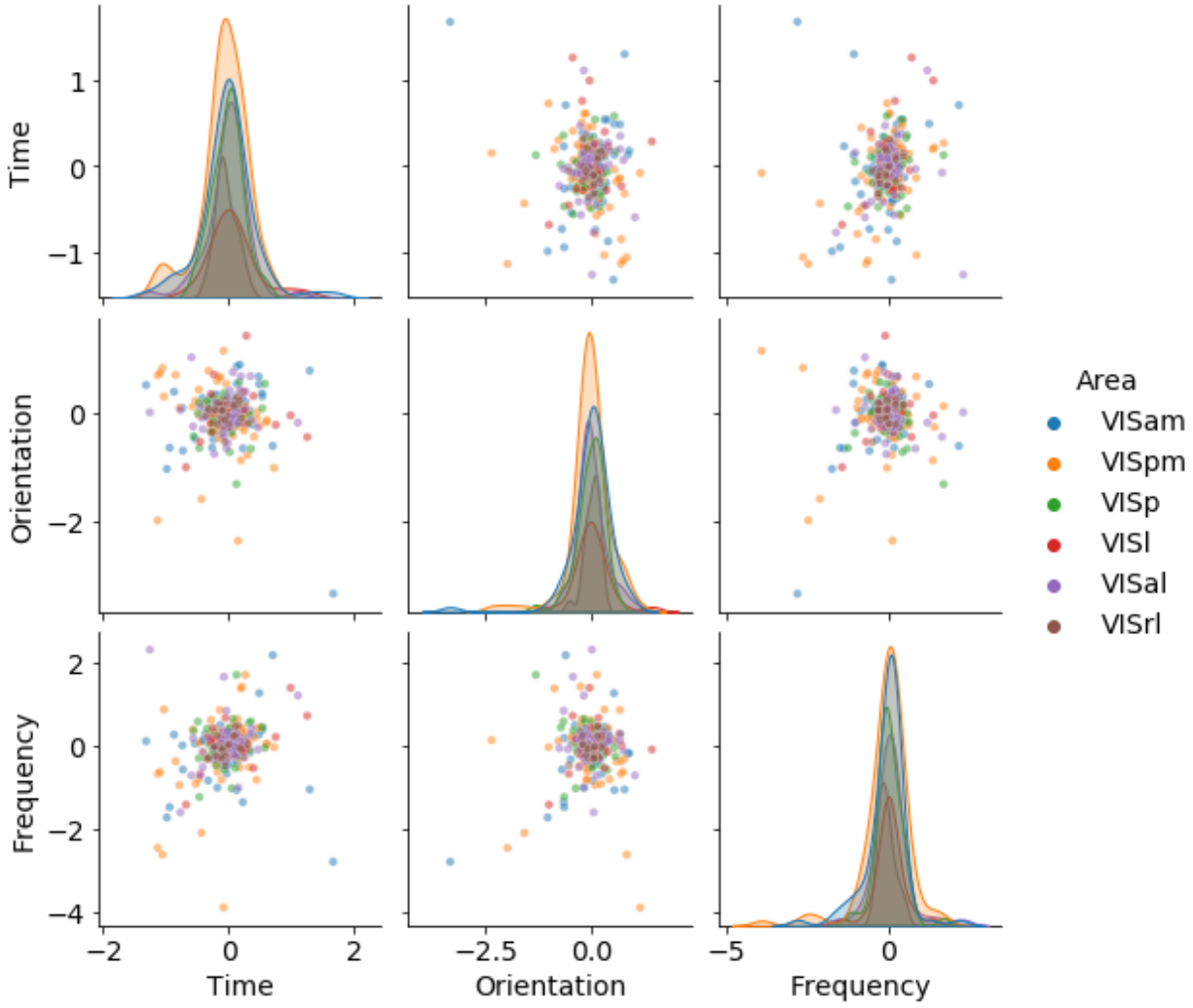

Figure 3. Distributions of dPCA weights for different component types across V1 structures.

When we inspect the contribution of each neuron to the first dimensions of dPCA projections, we don't find that the representations are dominated by any brain structure in particular within V1. Rather, the weights of dPCA seem to indicate that the representations that distinguish the stimuli are distributed across the visual cortex.

Consistent latent dynamics underlie V1 activity across animals

When we extracted the 10D latent dynamics of the visual cortex using PCA on different sessions, we were able to linearly align them using CCA and achieve high correlation. This shows that the structure of the latent dynamics of the visual cortex during drifting grating stimuli is similar across animals. For other areas, we were not able to achieve the same quality of alignment, showing that the activity of these functional areas is less consistent across animals.

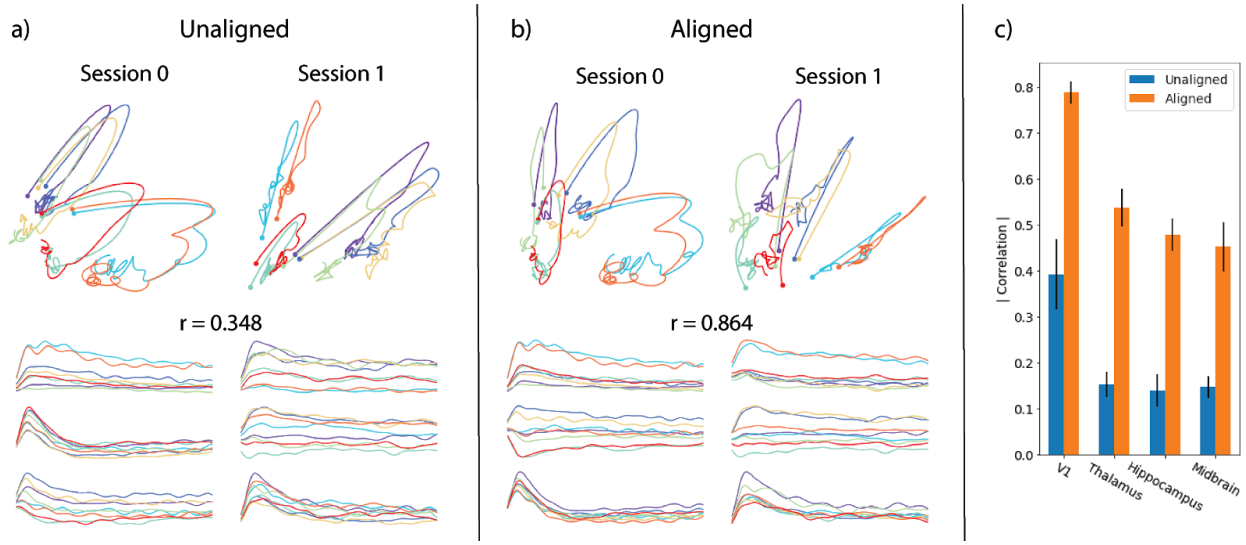

Figure 4. Aligning latent dynamics of PSTHs for 8 grating orientations (temporal frequency 4 Hz) across sessions. (a) The top 3 PCs plotted in 3D and as single-dimensions through time. Match between PSTHs is computed as the average absolute Pearson's r. (b) The 3D latent dynamics after alignment via PCA. (c) Alignment quality for different brain areas between all pairs of sessions. Error bars are standard deviations.

Population state in V1 areas is predictive of spikes in other V1 areas

As expected, the latent states best predicted spikes in the same area, but we had a few other fairly consistent findings. First, hippocampal areas CA1, CA3, and DG all predicted each other relatively well. Second, all of the visual cortex areas, denoted here by V1, typically predict each other relatively well. Overall, there was some consistent structure despite different sessions and different mice. However, there was quite a bit of variability as well, which we attribute to different distributions and numbers of recorded neurons and variability in attention during the movie.

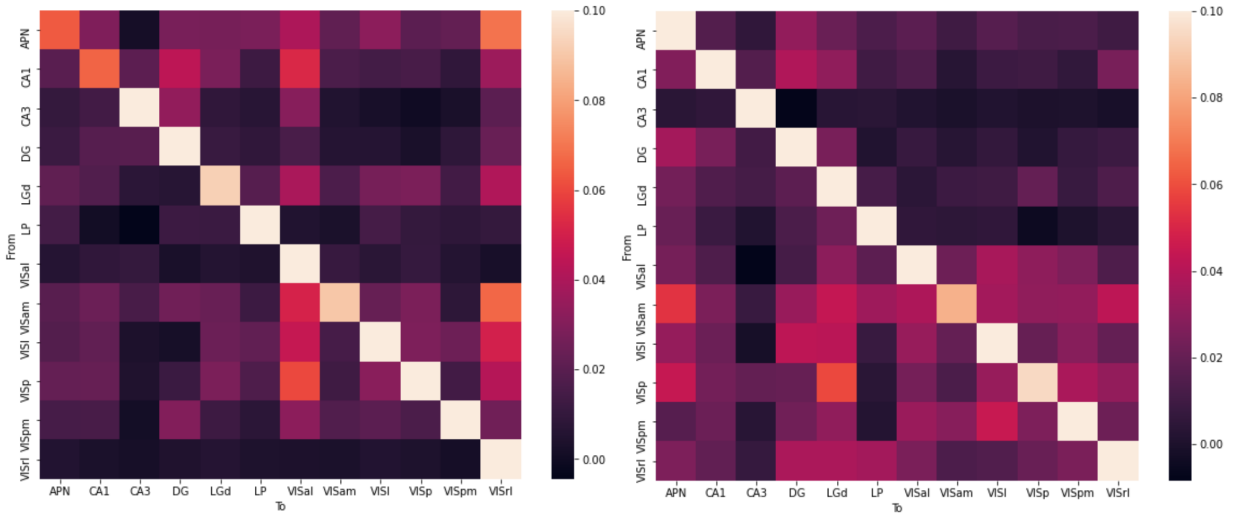

Figure 5. Heatmap of median pseudo-R^2 for predicting spikes in two example sessions (capped at 0.1 to reveal less obvious structure). Source areas on the left were the source of the predictive factors and target areas on the bottom were the sources of predicted spikes.

Natural Movies

Population state is still predictive of spikes across areas

Despite the increased complexity of the stimulus, similar results as drifting gratings were found when predicting spikes across brain areas within a session. We can see some interesting results from the heatmap, even if regression is not a very effective tool in projection. We can see that it is not necessarily symmetric, as projections from one area to another are not identical in the reverse, which makes sense as each region has different neuron population sizes. One can see that most regressions were able to map relatively well to APN and LGd. The regions in the visual cortex were also able to project to each other better than other regions were.

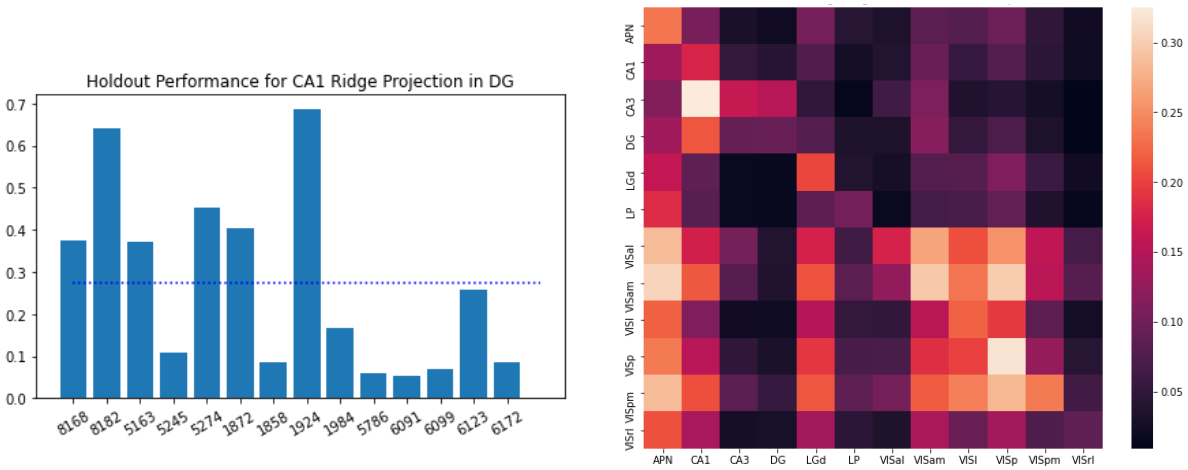

Figure 6. On the left is the R^2 between the actual spiking activity for each neuron in the DG area and the predicted spikes from a ridge regression derived projection of the CA1 brain area. The average is then represented in the heatmap on the right across the brain areas for a single session.

This predictive power can also be seen with example PSTH predictions below. While some general trends were predicted by the ridge regression, the Poisson regression applied to the latent factors most accurately modeled the spiking behavior. The model performance shows that the latent dimensions from simultaneous recordings of different brain areas have enough similar information embedded in them to map spiking activity relatively well. This is the case even with more dynamic stimuli such as the natural movies.

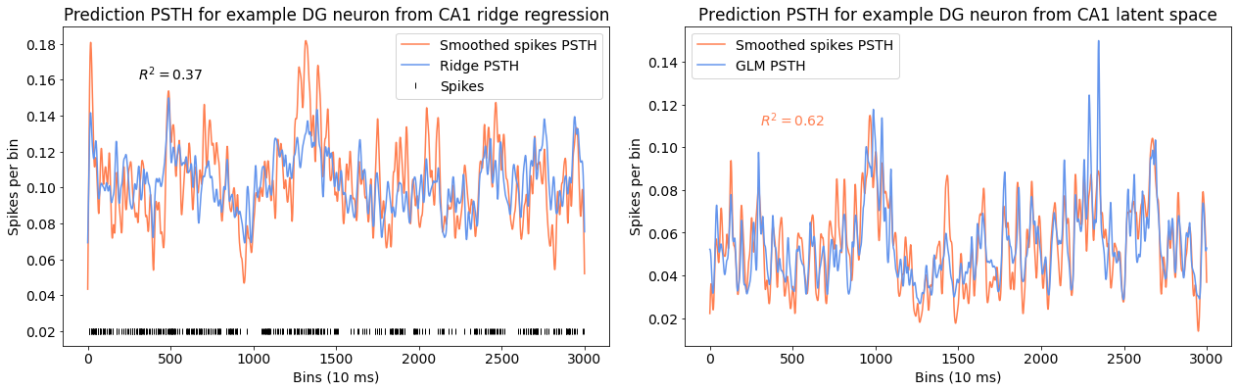

Figure 7. PSTH predictions from ridge regression on smoothed spikes (left) and poisson GLM on latent factors (right). Both capture the general trend, but the latent factors with the Poisson distribution more effectively model the spiking activity.

Population activity is alignable across sessions

In our first CCA experiment on the natural movie responses, we aligned data across all 12 brain structures of interest for five sessions. Each session contained a total of approximately 500 neurons, and we reduced this dimensionality to 3 in the CCA models. For all three sub-experiments, we found that the R^2 scores were very high (>0.8), even for the alignment of neural responses of different mice to different natural movies. For cases in sub-experiments 1 and 2 where we aligned the same neural response data to itself, the R^2 score was 1.0, as anticipated. However, we expected the results of the third sub-experiment where we aligned neural responses of different mice to different natural movies to have much lower R^2 scores. We believe that the alignment accuracy was high for all three sub-experiments because we reduced a very high-dimensional dataset to a latent dimensionality that was too low. Because of this, we chose to re-run this experiment using a higher latent dimensionality of 20.

Our second CCA experiment tested whether the R^2 scores of the CCA models from experiment 1 would drop significantly after increasing the latent dimensionality from 3 to 20. The latent dimensionality increase was the only change from experiment 1. The R^2 scores for this experiment were lower than those of experiment 1, with an average score of about 0.8. While this change is small, the decrease in R^2 scores after only increasing the latent dimensionality suggests that the density of the input data (i.e. the number of neuron responses being examined) may have an impact on the alignment results. To examine this further, we ran a final CCA experiment that only examined the neural responses from the primary visual area (VISp).

In our final CCA experiment we aligned neural responses from VISp for five sessions. Each session contained about 80 neurons, and we reduced this dimensionality to 3 in the CCA

models. For all alignments that did not align the same data with itself, the R^2 scores were much lower than those of experiments 1 and 2, with an average score of about 0.5. Alignments of the same data still produced R^2 scores of 1.0, as expected. The lower R^2 scores for this experiment also suggest that the density of input data impacts the quality of alignment, since this input data contained neural responses from about 80 neurons versus 500 in the previous experiments. The third sub-experiment also produced much lower R^2 scores than those of the previous experiments, with an average score of about 0.3, as we originally anticipated. The R^2 scores for all three experiments are summarized in Figure 8.

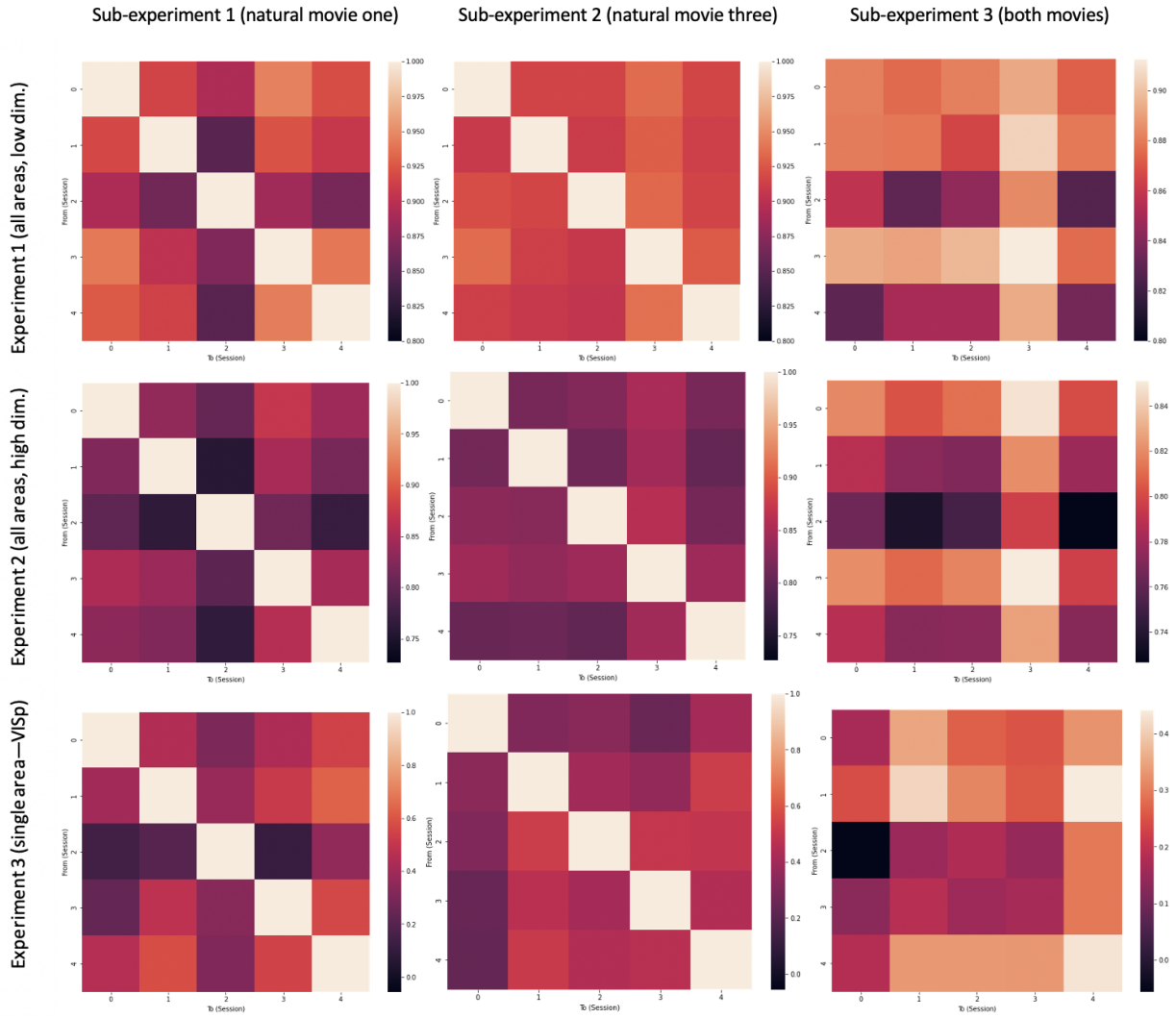

Figure 8. Heatmap of R^2 scores for aligning neural responses of different mice to natural movies. In sub-experiment 3, neural responses to natural movie one (represented by the y-axis) are aligned with neural responses to natural movie three (represented by the x-axis).

Conclusion

Stimulus Parameter Encodings within Responses to Drifting Gratings in V1

Our first results with the drifting grating stimuli showed clearly different patterns of population activity across the range of brain areas we analyzed. These results are reasonable due to the wide range of functions and neural firing characteristics across these brain structures. The only structure that showed a clear relationship to the orientation of the gratings was V1. This was made evident by the similarity and close proximity between latent trajectories during up-down grating conditions relative to all the other moving grating directions. Following this result, we used dPCA to explicitly extract visual stimulus parameter latents from the population activity. The results from this analysis showed clear separation across the different stimulus parameters, and even resulted in several highly interpretable 3D plots showing the low dimensional features that correlate most with each parameter of the visual stimuli. This information tended to be distributed evenly across the entire visual cortex, providing evidence against a simple modular representation of the parameters.

Once we had found evidence of the visual stimulus parameters being encoded in the latent dynamics of V1, we found that the structure of these encodings is common across mice, and could be highly correlated if properly aligned with CCA. As a final analysis, we explored the concept that these latent structures might be predictive across brain areas during the same stimuli. We found consistent evidence of predictability across brain areas in V1, which was consistent with our finding that the condition-specific encodings were distributed across the structure as a whole.

Alignment of Natural Movie Responses Suggests Common Dynamics

While our results in aligning neural responses to natural movies using CCA suggested that there exist some consistent latent dynamics within natural movie responses, more work needs to be done to conclude this. The method of using PCA to reduce dimensionality and then applying CCA to the latent-dimensional data could be done to further this argument, as we did for neural responses to drifting gratings. We would also need to examine the other 30 second increments of natural movie three to better understand any shared dynamics between natural movies taken from the same longer film. Reducing dimensionality and aligning neural responses to natural movies not taken from the same film as natural movies one and three would also help to establish whether there are common latent dynamics within natural movies of specific settings, movements, etc.

In Summary

Our results exhibit the diverse range of dimensionality reduction techniques that can be applied to uncover specific aspects of neural processing through analysis of the latent features. From our analyses, we were able to discern how contributions from hundreds of individual neurons in the mouse visual cortex can be coordinated in structured ways to process features from the

external environment. These structures were so robust that we were even able to show similarities in latent features across animals. Because these latent features evolve dynamically as the stimulus is shown, we believe they reflect an underlying functional process, rather than simple representation.

References

[1] M. Churchland *et al.*, “Neural population dynamics during reaching,” *Physiol. Behav.*, vol. 176, no. 3, pp. 139–148, 2019.

[2] J. E. K. Miller, I. Ayzenshtat, L. Carrillo-Reid, and R. Yuste, “Visual stimuli recruit intrinsically generated cortical ensembles,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 38, pp. E4053–E4061, 2014.

[3] J. Pérez-Ortega, T. Alejandre-García, and R. Yuste, “Long-term stability of neuronal ensembles in mouse visual cortex,” *bioRxiv*, p. 1, 2020.

[4] S. Vyas, M. D. Golub, D. Sussillo, and K. V. Shenoy, “Computation through Neural Population Dynamics,” *Annu. Rev. Neurosci.*, vol. 43, pp. 249–275, 2020.

[5] E. Froudarakis *et al.*, “Object manifold geometry across the mouse cortical visual hierarchy,” *bioRxiv*, p. 2020.08.20.258798, 2020.

[6] G. Santhanam *et al.*, “Factor-analysis methods for higher-performance neural prostheses,” *J. Neurophysiol.*, vol. 102, no. 2, pp. 1315–1330, 2009.

[7] S. E. J. de Vries *et al.*, “A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex,” *Nat. Neurosci.*, vol. 23, no. 1, pp. 138–151, 2020.

[8] “Allen Brain Observatory : Visual Coding Neuropixels Dataset,” vol. 0, no. October, 2019.

[9] J. A. Gallego, M. G. Perich, R. H. Chowdhury, S. A. Solla, and L. E. Miller, “Long-term stability of cortical population dynamics underlying consistent behavior,” *Nat. Neurosci.*, vol. 23, no. 2, pp. 260–270, 2020.

[10] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011.

[11] M. Tolkiehn and S. R. Schultz, “Temporo-nasally biased moving grating selectivity in mouse primary visual cortex,” *bioRxiv*. bioRxiv, p. 708644, 19-Jul-2019.