Automatic Alignment of Neural Data by Piecewise Linear Time Warping

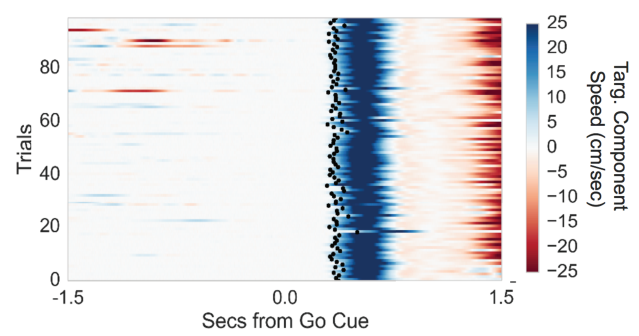
Alex H. Williams[1], Ben Poole[2], Niru Maheswaranathan[1], Ashesh K. Dhawale[10,11], Tucker Fisher[1], Christopher D. Wilson[14], David H. Brann[10,11], Eric Trautmann[1], Stephen Ryu[12], Roman Shusterman[13], Dmitry Rinberg[14,15], Bence P. Olveczky[10,11], Krishna V. Shenoy[4,5,6,7,8,9], Surya Ganguli[3,4]

Summary. We developed an unsupervised, data-driven method, *Piecewise Linear Warping*, to model systematic variability in spike timing. This method aligns high-dimensional neural spike trains across trials and discovers surprising structure that is rendered invisible by naïve trial-averaging.

Trial-to-trial variability in timing is ubiquitous

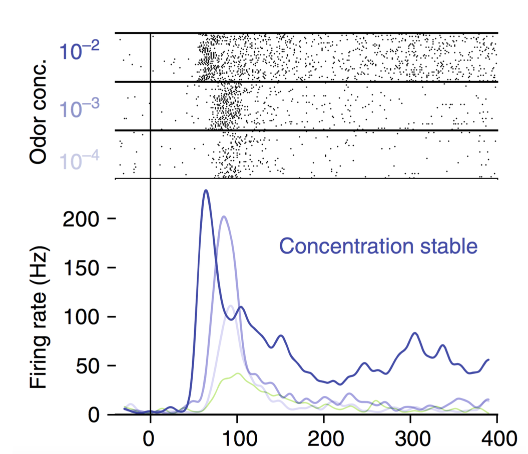
Nuisance variation in reaction times

In simple cued reaching behaviors, non-human primates show variable reaction times on the order of several hundred milliseconds (Khanna, 2017). Should one align each trial to the stimulus presentation ("Go Cue") or movement onset or maximum hand velocity?



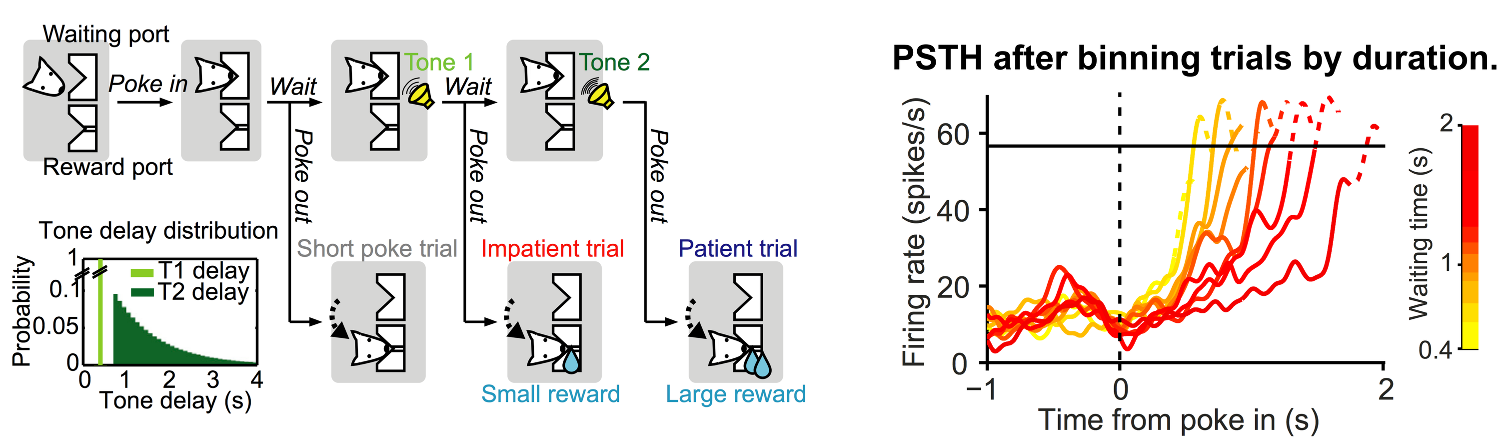
Stimulus-driven variation in timing

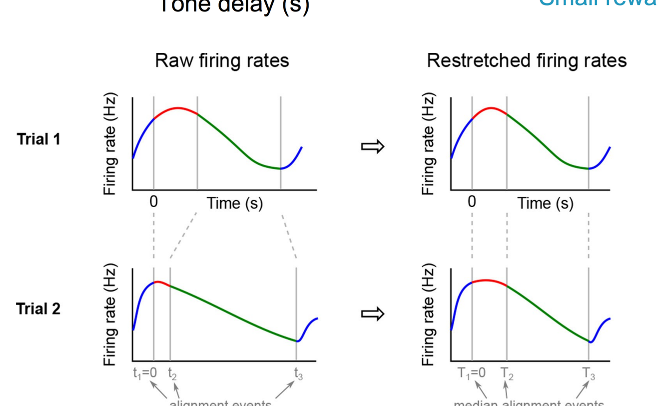
In many systems, a stronger stimulus will elicit a faster neural or behavioral response. In olfaction, higher concentrations of an odorant evoke earlier responses in mitral-tufted cells; these temporal effects may provide a mechanism for encoding odor identity (Wilson et al., 2017). How can we detect similar effects in other systems, where the stimulus is not as precisely controlled?



Response-driven variation

In self-paced experiments, animals can end trials early leading to highly variable trial durations. An interesting idea is to average neural spike trains over trials of similar duration (right; Murakami et al. 2014), but can we develop more principled analyses that capture single-trial variations in timing across full populations of neurons?

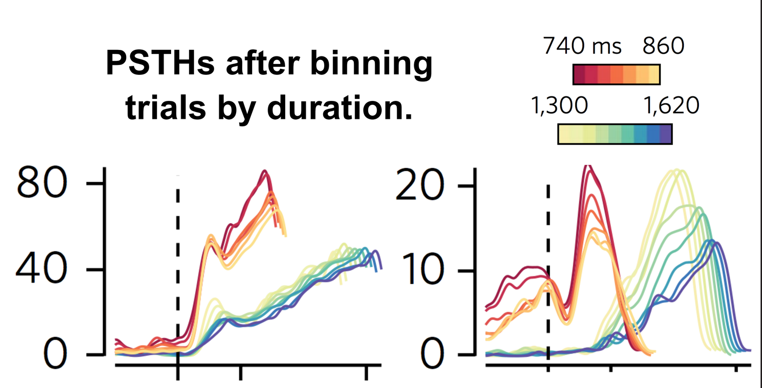




Another idea is to define multiple alignment events for each trial (e.g. nose poke times, stimulus presentations, etc.) and then linearly warp each segment of every trial to the median time of each alignment point (left; Kobak et al., 2016). Are human-annotated alignment points optimal, or can a statistical model identify better alignments?

Computation-driven variation

Finally, it has been proposed that neural dynamics are stretched/compressed during production of timed motor sequences (Wang et al., 2018). Statistical models of systematic temporal variability across neural populations may give direct insight into computational mechanisms.



[1]Neuroscience Program, [2]Computer Science Department, [3]Applied Physics Department, [4]Neurobiology Department, [5]Electrical Engineering Department, [6]Bioengineering Department, [7]Bio-X Program, [8]Stanford Neurosciences Institute, [9]Howard Hughes Medical Institute. Stanford University, Stanford, CA 94305, USA. [10]Department of Organismic and Evolutionary Biology, [11]Center for Brain Science Harvard University, Cambridge, MA 02138, USA. [12]Department of Neurosurgery. Palo Alto Medical Foundation, Palo Alto, CA 94301, USA.[13]Institute of Neuroscience University of Oregon, Eugene, OR 97403, USA. [14]NYU Neuroscience Institute, [15]Center for Neural Science New York University Langone Medical Center, New York, NY 10016, USA.

This work was supported by the U.S. Department of Energy Computational Science Graduate Fellowship (CSGF) program, NIH NRSA 1F31NS089376-01, Burroughs Wellcome Foundation Sloan Foundation, Simons Foundation, Office of Naval Research McKnight Foundation.

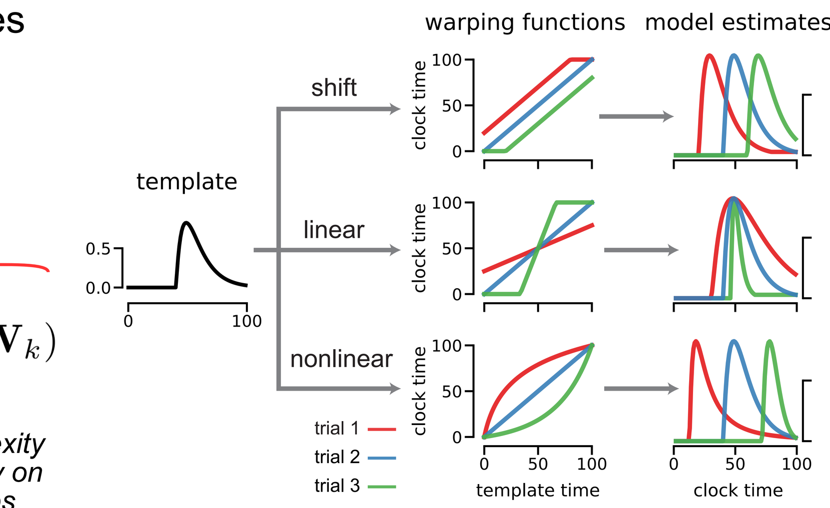
Time warping models capture timing variability

The key idea is to fit a template time series that is shifted and/or stretched—i.e., *warped*—on a trial-by-trial basis to match the data. Formally, the model minimizes:

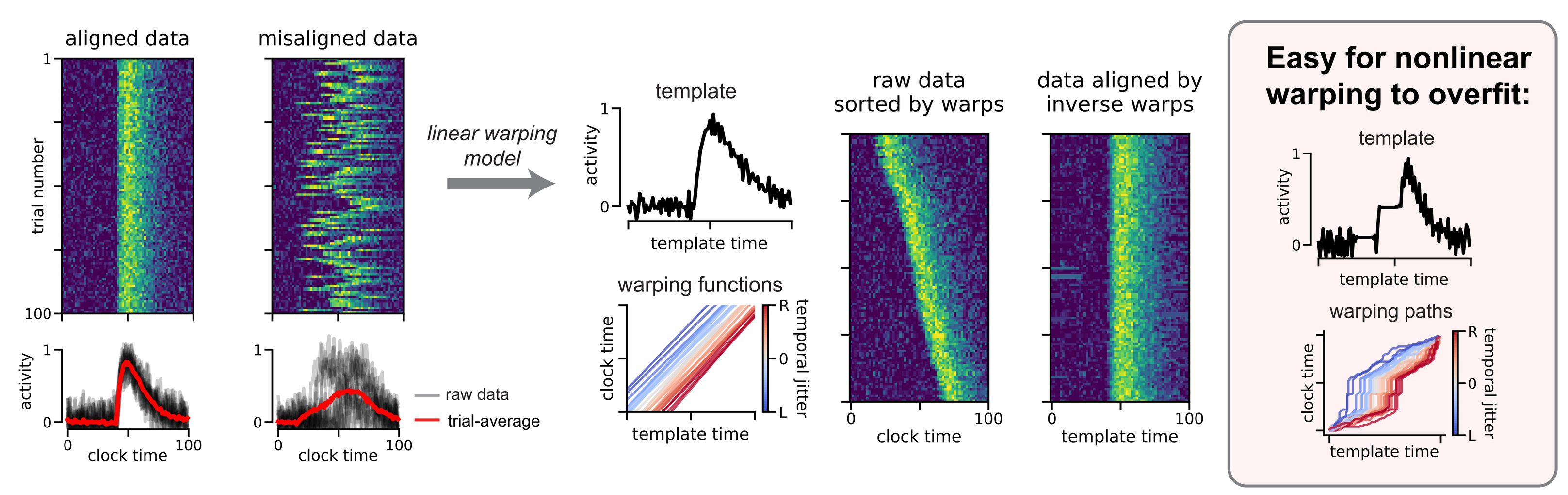
$$\underbrace{\sum_{k=1}^{K} \left\| \mathbf{W}_k \widetilde{\mathbf{X}} - \mathbf{X}_k \right\|_F^2}_{\text{reconstruction error*}} + \underbrace{\rho_1(\widetilde{\mathbf{X}}) + \sum_{k=1}^{K} \rho_2(\mathbf{W}_k)}_{\text{regularization}}$$

warp on trial k; template; data on trial k; roughness penalty on template; complexity penalty on warps

*In addition to least-squares loss, our code also supports Poisson loss.

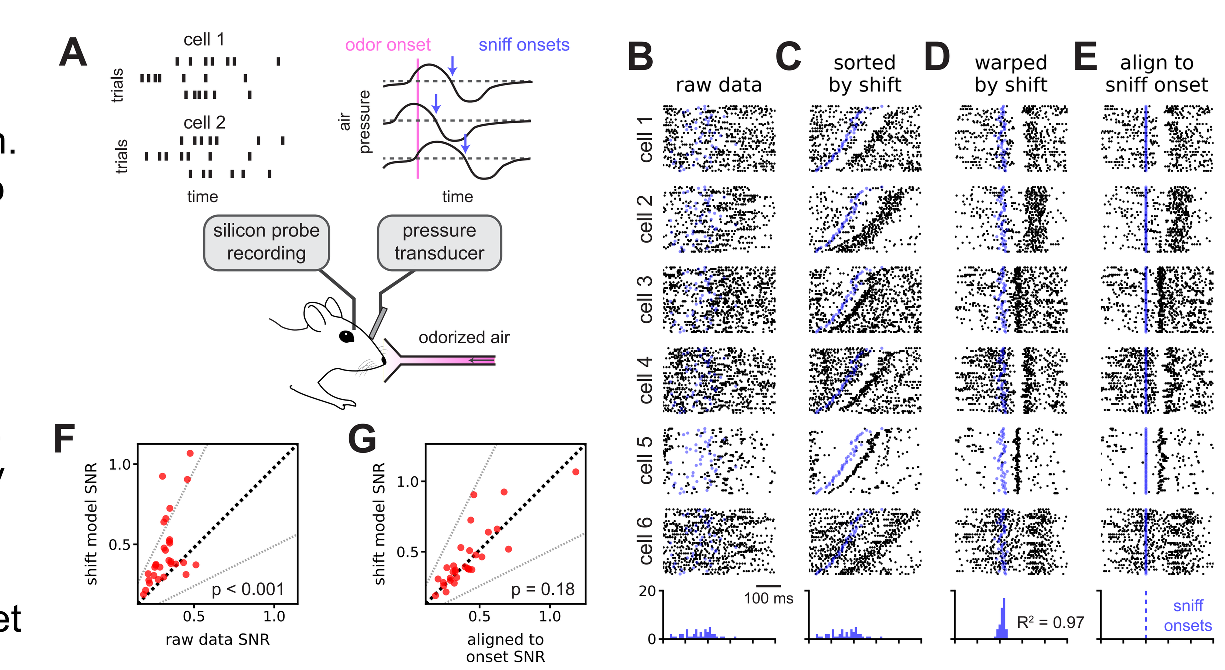


Example of linear warping on one neuron (synthetic data):

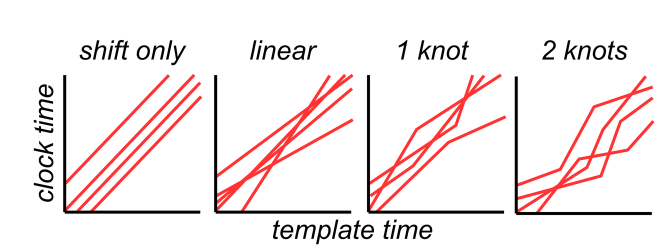


Case Study #1: Mouse olfaction

We fit the model to spike trains aligned to odor presentation. The model learns to align responses tightly with sniff onset (a more proximal event to receptor activation). The model arguably provides a *better* alignment than measured sniff onset (see e.g., cell #4).



Some advantages of our approach:



Simplicity. We developed methods to fit models with piecewise-linear warping functions. This allows us to gradually increase the complexity of warping.

Few assumptions. Previous work has assumed that neural activations are low-dimensional; we make no assumption on the dimensionality of dynamics (and have found it harmful in practice!)

Computationally scalable. Datasets involving hundreds of neurons and trials can typically be fit in less than a minute on a modern laptop. We accomplished this by deriving a closed-form update rule for the template. The warping functions for each trial are fit by an unbiased random search in parallel across CPU threads.

Related work:

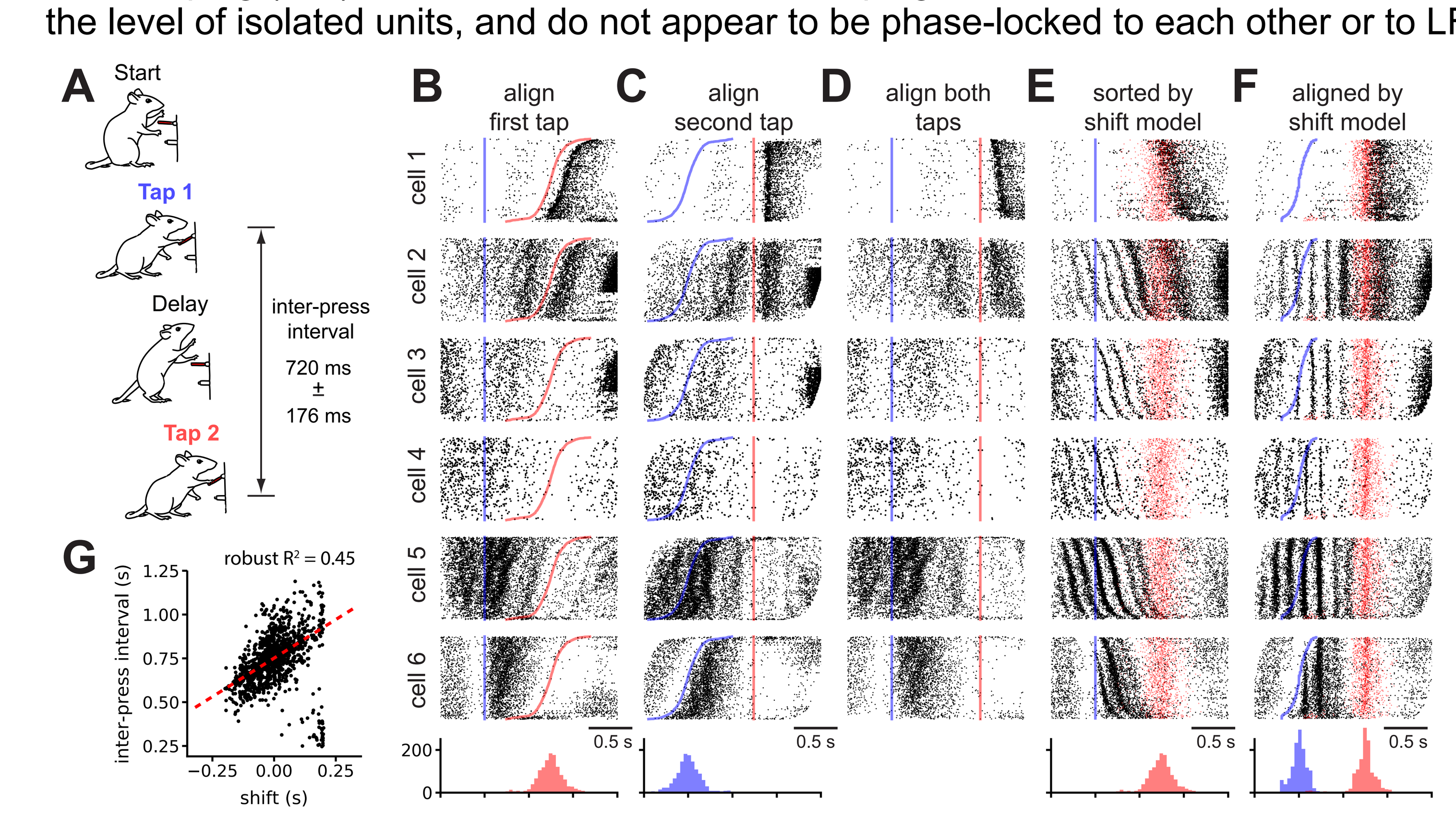
Time-warped PCA *(Poole et al., Cosyne 2016).* Previous work by our group. Parameterizes warping functions as a dense grid of knots and uses gradient-based optimization. Assumes low-dimensional dynamics. More computationally intensive; can get caught in local minima.

Time-warped GLMs *(Lawlor et al., J Comput Neurosci; 2018).* Builds time warping into regression/supervised methods. Uses nonlinear warping functions.

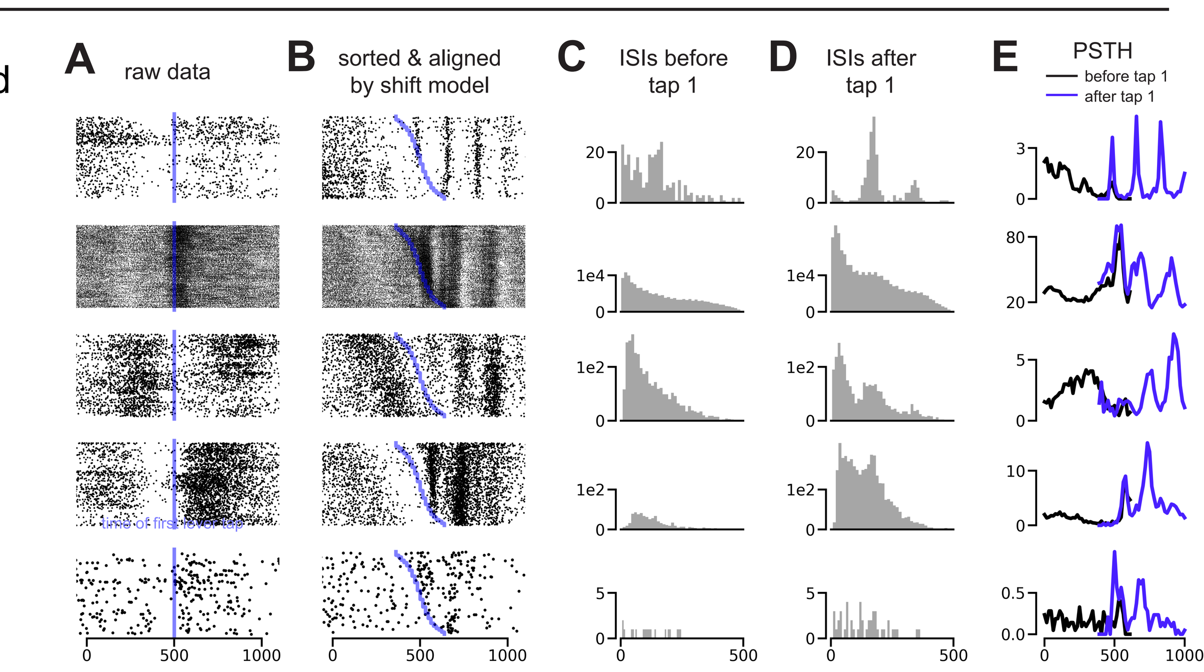
Gaussian Process Alignment *(Duncker & Sahani, bioRxiv 331751; 2018).* Parameterizes warping functions as Gaussian processes; nonlinear and not necessarily monotonically increasing. Fully probabilistic model tailored to Poisson distributed spike times. Assumes low-dimensional dynamics.

Case Study #2: Rat motor cortex

A rat taps a lever twice with a target wait time. Aligning each trial to one or both lever presses (B-D) obscures striking theta-band oscillations, which are revealed by shift-only time warping (E-F) without need for nonlinear warping. These oscillations are visible at the level of isolated units, and do not appear to be phase-locked to each other or to LFP.

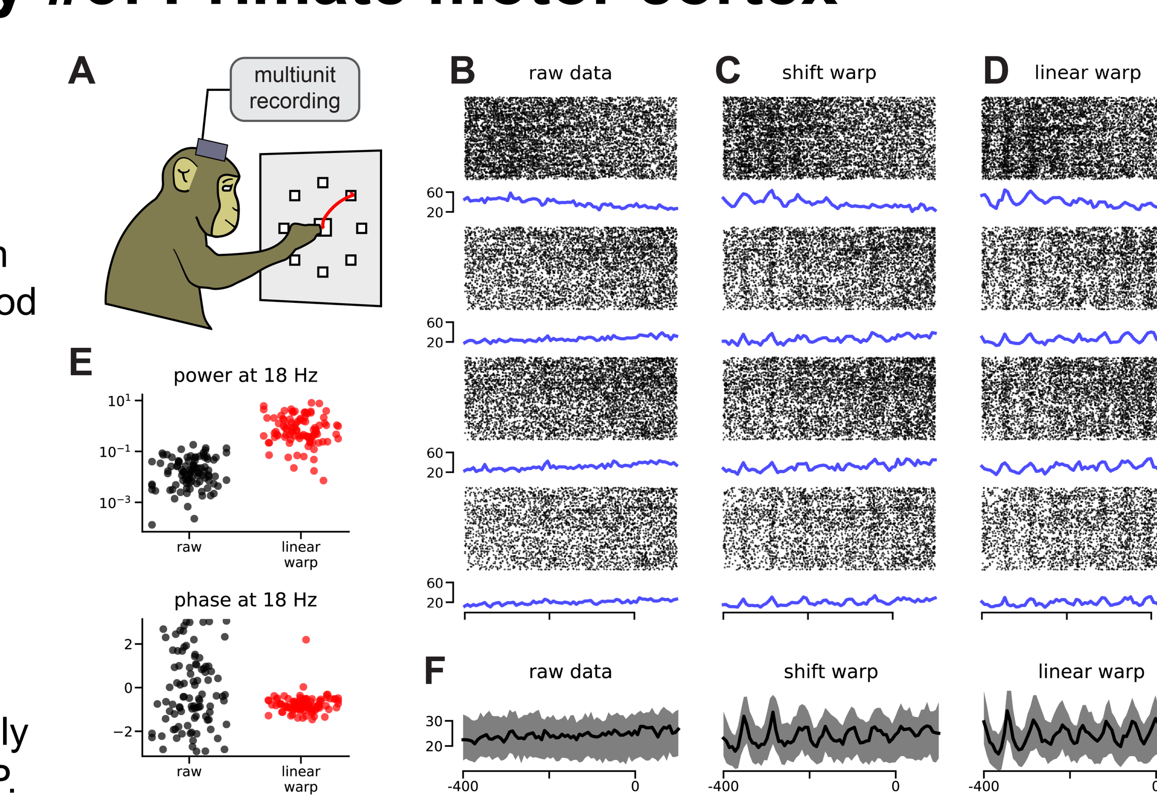


Zooming in around the first lever press (blue) shows that spike oscillations are initiated precisely after the lever press event (not by movement onset). This effect is not seen in all neurons.



Case Study #3: Primate motor cortex

A Rhesus monkey made cued radial reaches in a standard motor assay. Focusing on the prepartory period (prior to "Go Cue") we find robust beta-band oscillations in multi-unit activity. Unlike case study #2, oscillations are in-phase across electrodes and likely correlated with LFP.



Similar oscillations were recently found by LFADS, a deep recurrent neural architecture, with single trial/unit resolution (Pandarinath et al., 2018). Linear time warping recapitulates this result with a simpler statistical model.

Conclusions. Remarkably simple time warping models can uncover striking dynamics that are invisible in raw data, even in brain areas relatively close to the sensory-motor periphery and in experimental tasks with well-defined alignment points. This method enables data-driven discovery of precise spike patterns that are likely overlooked by any trial-averaged analysis.