

# Discovering precise temporal patterns in large-scale neural recordings through robust and interpretable time warping

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

<sup>1</sup>Neuroscience Program, <sup>2</sup>Computer Science Department, <sup>3</sup>Applied Physics Department, <sup>4</sup>Neurobiology Department, <sup>5</sup>Electrical Engineering Department, <sup>6</sup>Bioengineering Department, <sup>7</sup>Bio-X Program, <sup>8</sup>Wu Tsai Stanford Neurosciences Institute, <sup>9</sup>Howard Hughes Medical Institute  
Stanford University, Stanford, CA 94305, USA.

<sup>10</sup>Department of Organismic and Evolutionary Biology, <sup>11</sup>Center for Brain Science  
Harvard University, Cambridge, MA 02138, USA.

<sup>12</sup>Department of Neurosurgery  
Palo Alto Medical Foundation, Palo Alto, CA 94301, USA.

<sup>13</sup>Institute of Neuroscience  
University of Oregon, Eugene, OR 97403, USA.

<sup>14</sup>NYU Neuroscience Institute, <sup>15</sup>Center for Neural Science  
New York University Langone Medical Center, New York, NY 10016, USA.

\*Send correspondence to [ahwillia@stanford.edu](mailto:ahwillia@stanford.edu), [sganguli@stanford.edu](mailto:sganguli@stanford.edu)

## 1 Abstract

2 Though the temporal precision of neural computation has been studied intensively, a data-driven determination  
3 of this precision remains a fundamental challenge. Reproducible spike time patterns may be obscured on single  
4 trials by uncontrolled temporal variability in behavior and cognition, or may not even be time locked to measurable  
5 signatures in either behavior or local field potentials (LFP). To overcome these challenges, we describe a general-  
6 purpose time warping framework that reveals precise spike-time patterns in an unsupervised manner, even when  
7 spiking is decoupled from behavior or is temporally stretched across single trials. We demonstrate this method  
8 across diverse systems: cued reaching in nonhuman primates, motor sequence production in rats, and olfaction in  
9 mice. This approach flexibly uncovers diverse dynamical firing patterns, including pulsatile responses to behavioral  
10 events, LFP-aligned oscillatory spiking, and even unanticipated patterns, like 7 Hz oscillations in rat motor cortex  
11 that are not time-locked to measured behaviors or LFP.

## 12 Introduction

13 The role of spike time precision in neural computation has been widely examined from both experimental and  
14 theoretical perspectives (Softky and Koch 1993; London et al. 2010; Bruno 2011; Amarasingham et al. 2012;  
15 Amarasingham et al. 2015; Brette 2015; Denève and Machens 2016), engendering intense debates in systems  
16 neuroscience over the last several decades. Empirically determining the degree of temporal precision from data is  
17 challenging because multi-neuronal spike trains may contain highly structured temporal patterns that are completely

18 masked by temporal variations in behavioral and cognitive variables not under direct experimental control. For  
19 example, precise spike patterns may not be temporally locked to naïvely chosen sensory or behavioral events.  
20 Indeed, the fidelity of olfactory coding may be underestimated by factors of two to four when spike times are aligned  
21 to stimulus delivery instead of inhalation onset (Shusterman et al. 2011; Cury and Uchida 2010; Shusterman et al.  
22 2018).

23 Thus, experimental estimates of spike time precision hinge on the choice of an alignment point, which defines the  
24 origin of the time axis on each trial. This choice can often be challenging and subjective. Even in relatively simple  
25 behavioral tasks, animals can experience a sequence of stimuli, actions, and rewards, each of which occur with  
26 varying latencies on different trials. Such tasks thus provide multiple choices for aligning multineuronal spike trains  
27 to measurable events marking an origin of time. Moreover, in addition to choosing an origin of time, we must also  
28 choose its units. Should spike times be measured in absolute clock time relative to some measured event, or in  
29 units of fractional time between two events? Should the units of time change between successive pairs of events?  
30 Could any one of these choices unmask spike-timing precision that is otherwise invisible?

31 Past studies have addressed these challenges in a number of ways: grouping trials together with similar durations  
32 before averaging spike counts (Murakami et al. 2014; Starkweather et al. 2017; Wang et al. 2018), manually  
33 stretching or compressing time units between measured task events (Leonardo and Fee 2005; Shusterman et al.  
34 2011; Kobak et al. 2016; Aronov et al. 2017), or repeating statistical analyses around different choices of alignment  
35 point (Feierstein et al. 2006; Harvey et al. 2012; Jazayeri and Shadlen 2015; Shushruth et al. 2018). Determining  
36 an appropriate alignment and scaling of time is most challenging in systems far from the sensory or motor periphery,  
37 where neural responses are often not locked to any external event, and instead reflect internal decisions or changes-  
38 of-mind that occur at variable times within each trial. In these cases, the ideal temporal alignment point (e.g. the  
39 time a decision is made) may be entirely unmeasurable or ill-defined from the standpoint of behavior.

40 These complications, and the diversity of heuristic approaches used to address them, underscore a broad need  
41 for statistical frameworks to assess the temporal precision of neural computation. Of particular interest are  
42 *unsupervised* statistical methods that reveal precise patterns in multi-neuronal spike trains *without* reference to  
43 behavioral measurements. Such methods would be broadly applicable, as they make few assumptions about  
44 experimental design, animal model, or measured behaviors. Furthermore, by only considering the neural data, these  
45 methods may discover novel spike patterns aligned to unexpected variables. Most intriguingly, these methods could  
46 potentially identify spike patterns that are not well-aligned to any behavior, but instead to unobservable cognitive  
47 states, such as decision times.

48 While time series and image alignment methods are a well-studied topic in signal processing (Berndt and Clifford  
49 1994; Marron et al. 2015; Mueen and Keogh 2016; Pnevmatikakis and Giovannucci 2017), these techniques  
50 have rarely been applied to large-scale neural recordings (but see recent work by Poole et al. 2017; Lawlor et al.  
51 2018; Duncker and Sahani 2018). Neuroscientists have historically utilized simple alignment operations—namely,  
52 translating (Baker and Gerstein 2001; Ventura 2004) and potentially stretching/compressing activity traces between  
53 pairs of behavioral events (Shusterman et al. 2011; Leonardo and Fee 2005; Perez et al. 2013; Kobak et al. 2016).  
54 In contrast, popular statistical methods, such as *Dynamic Time Warping* (DTW; Berndt and Clifford 1994), allow  
55 signals to be non-uniformly compressed and dilated on each trial. While such *nonlinear warping* models can be  
56 useful, we demonstrate that they can be difficult to interpret and sensitive to the high level of noise that is typical of  
57 neural data.

58 To identify interpretable alignments for high-dimensional spike trains, we developed a framework for *linear* and  
59 *piecewise linear* time warping that encompass existing human-annotated procedures (Leonardo and Fee 2005;  
60 Kobak et al. 2016). Relative to nonlinear warping methods, the methods we describe are simpler to interpret, more  
61 robust to overfitting, and more computationally scalable. We applied these methods to multielectrode recordings

62 collected from three experiments spanning different animal models (rodents and primates), brain regions (olfactory  
63 and motor cortex), and behavioral tasks (sensation and motor production). In each case, time warping revealed  
64 precise spike patterns that were imperceptible in the raw data. Moreover, some of these results were *not* easily  
65 captured by commonly chosen temporal alignments. For example, in rodents performing a motor timing task,  
66 we uncovered prominent ~7 Hz oscillations in spike times that were *not* aligned to the LFP or any measured  
67 behavioral event, providing a convincing example in which salient population dynamics would be likely overlooked  
68 by pre-defined behavioral alignments.

69 Overall, we demonstrate that simple time warping models can detect salient, yet easy-to-overlook, features in  
70 large-scale neural data. We expect these methods to particularly enable inquiry into circuits far from the sensory  
71 or motor periphery that are not tightly time-locked to any measurable stimulus or behavior. Yet, even in cases  
72 with seemingly obvious sensory or behavioral alignment points, our immediate findings suggest that common  
73 manual alignment and trial-averaging procedures may underestimate temporal precision, and may even qualitatively  
74 misrepresent single-trial dynamics. Thus, we provide a broadly applicable, data-driven framework to reveal and  
75 scientifically characterize fine-scale temporal features of neural dynamics. Such a framework, when combined with  
76 future experiments, may help in adjudicating long-standing debates surrounding the temporal precision of neural  
77 computation.

## 78 Results

### 79 Time Warping Framework

80 Our ultimate goal is to identify dynamical firing patterns that reliably occur on a trial-by-trial basis. If these activity  
81 patterns are tightly time-locked to a sensory or behavioral event, which can be confidently identified and measured,  
82 then we can accurately characterize the neural response by averaging over trials. This is illustrated in **Figure 1a**  
83 (left) which shows 100 trials of a simulated neural activity trace with additive Gaussian noise. The average activity  
84 trace (red trace; bottom) extracts the reliable features of the neural response from noisy single-trial instantiations  
85 (semi-transparent black traces). This synthetic example loosely resembles calcium fluorescence traces, but the  
86 methods we describe can be flexibly applied to any multi-dimensional time series including spike trains, fMRI data,  
87 or LFP traces.

88 More formally, if  $N$  neurons are measured at  $T$  timepoints over  $K$  nominally identical trials, the trial-average is given  
89 by:

$$\bar{\mathbf{X}} = \frac{1}{K} \sum_{k=1}^K \mathbf{X}_k. \quad (1)$$

90 Here,  $\mathbf{X}_k$  is an  $N \times T$  matrix denoting the measured activity of all neurons on trial  $k$ . In the context of spiking data,  
91 each row of  $\bar{\mathbf{X}}$  corresponds to a *peri-stimulus time histogram* (PSTH) of a recorded neuron. Trial averaging is also a  
92 common step in population-level statistical analyses (Gao and Ganguli 2015; Kobak et al. 2016).

93 Despite its widespread use, trial averaging can produce highly inaccurate and misleading results when neural  
94 activity is misaligned across trials. For example, introducing a random temporal shift to each simulated trial results  
95 in a less informative trial-average trace (**Fig 1a**, right). Such jitter commonly arises in practical applications, leading  
96 many research groups to develop custom-built alignment procedures for their experimental system. For example,  
97 in songbirds it is common to manually segment and cluster song syllables and warp intervening spike times on a  
98 per-syllable basis (Leonardo and Fee 2005). In olfaction, detailed measurements and fluid dynamics modeling  
99 of the sniff cycle have been pursued to understand the accuracy of sensory responses (Shusterman et al. 2011;

100 Shusterman et al. 2018). Jitter and other forms of temporal variability are likely exacerbated in deeper brain areas,  
101 where there is greater opportunity for unobserved cognitive states (e.g. attentiveness) and actions (e.g. internal  
102 decision times) to influence the timing of dynamics.

103 Time warping methods address these challenges through a data-driven, statistical approach. The key idea is to fit a  
104 *response template* time series that is shifted and stretched—i.e., *warped*—on a trial-by-trial basis to match the data.  
105 The response template, denoted  $\tilde{X}$ , is an  $N \times T$  matrix of activity traces that captures the average activity across  
106 trials *after* correcting for temporal deformations.

107 The time axis of the response template is then transformed by a *warping function* on each trial. Formally, we  
108 denote the warping function for trial  $k$  as  $\omega_k(t)$ ; this function maps each timebin  $t$  (clock time) to a new index  $\omega_k(t)$   
109 (template time). If  $\omega_k(t)$  is an integer between 1 and  $T$ , then the warping transformation for every neuron  $n$  on trial  $k$   
110 amounts to the transformation  $\tilde{X}_{nt} \mapsto \tilde{X}_{n\omega_k(t)}$ . If  $\omega_k(t)$  is *not* an integer then time warping is implemented by linear  
111 interpolation (see *Methods*). Note that this model assumes all recorded neurons share the same warping function  
112 on a trial-by-trial basis, though this could be relaxed by future work.

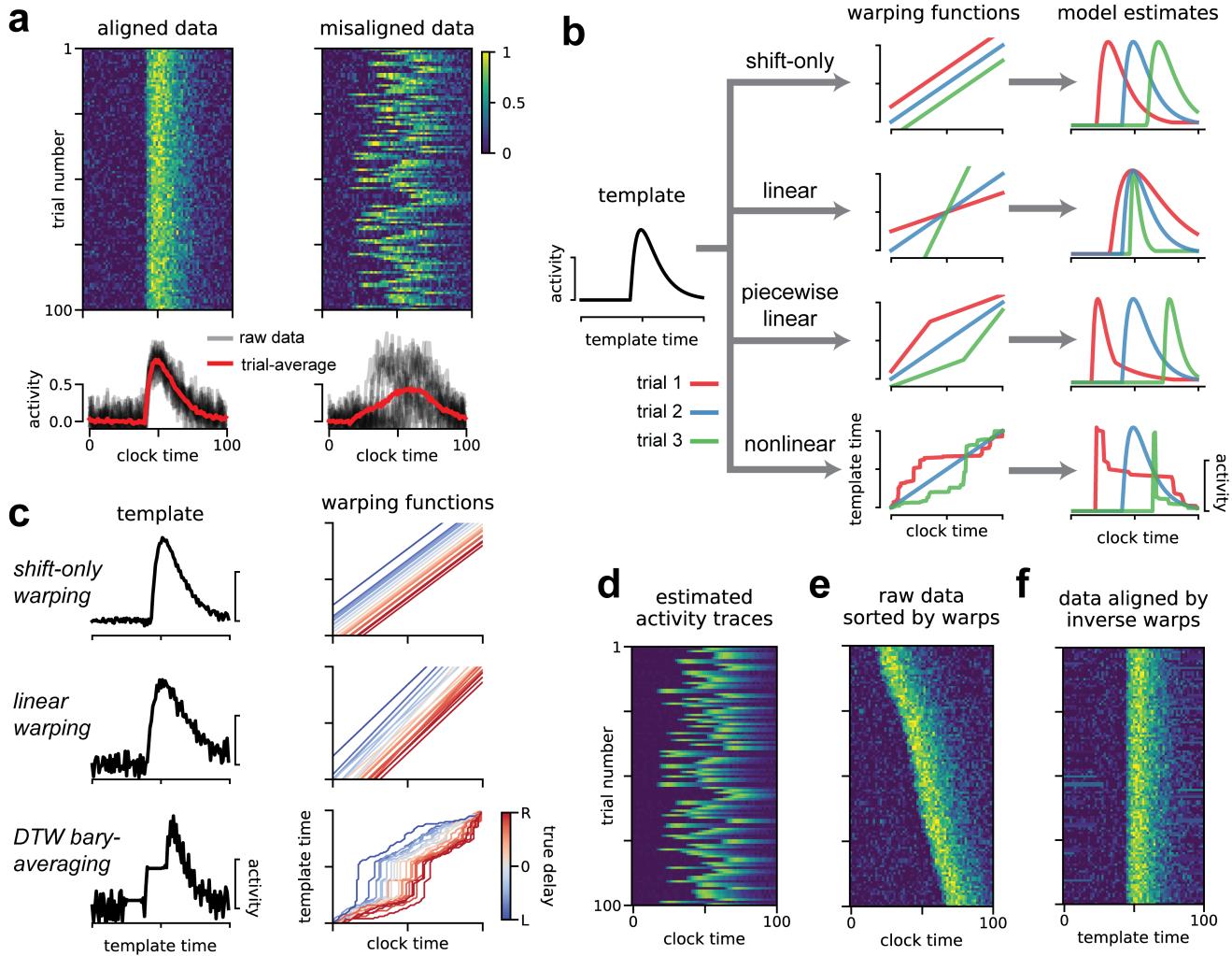
113 **Figure 1b** illustrates how different classes of warping functions account for single-trial variability in timing. In this  
114 paper, we focus on three main model classes: *shift-only time warping*, *linear time warping*, and *piecewise linear time*  
115 *warping* (**Fig 1b**, top three models). Shift-only warping represents the simplest model: the warping functions are  
116 constrained to be linear with slope equal to one, and only a single parameter (the y-intercept of  $\omega_k(t)$ ) is fit on each  
117 trial. As its name suggests, the shift-only model can only account for trial-to-trial differences in response latency. In  
118 contrast, a linear warping model, which fits the slope in addition to the intercept of  $\omega_k(t)$ , can account for variable  
119 latencies as well as uniform stretches or compressions of the response template. A piecewise linear warping model  
120 adds further complexity by adding one or more *knots* (points where the slope of  $\omega_k(t)$  can change). Most generally,  
121 nonlinear warping functions may be used, which non-uniformly stretch and compress portions of the template on  
122 each trial (**Fig 1b**, bottom).

123 In all cases, we constrain the time warping functions to be monotonically increasing. Intuitively, this ensures that the  
124 model cannot go backwards in time while making a prediction—that is, as  $t$  (clock time) increases,  $\omega_k(t)$  (template  
125 time) must also increase. This ensures that the warping functions are invertible, which we later exploit to align data  
126 across trials. DTW-based time warping paths are not invertible, since the first derivative can be zero or infinite.  
127 Some other nonlinear warping methods (e.g. Duncker and Sahani 2018) do not require warping functions to be  
128 monotonic (and therefore invertible), though future work on these methods could incorporate this constraint.

129 The warping functions and response template are numerically optimized to minimize the total reconstruction error  
130 over all neurons, trials, and timepoints. For flexibility and computational efficiency, we chose the mean squared  
131 error to quantify reconstruction accuracy. Ignoring the interpolation step of time warping for the sake of clarity (see  
132 *Methods*), the total model error is:

$$\frac{1}{NTK} \sum_{n=1}^N \sum_{t=1}^T \sum_{k=1}^K (\tilde{X}_{n\omega_k(t)} - X_{ntk})^2 \quad (2)$$

133 This expression is minimized with respect to the warping functions on each trial,  $\omega_k(t)$ , and the response template,  
134  $\tilde{X}$ . Other loss functions may be substituted for the mean squared error. In particular, a loss function based on  
135 Poisson noise is popular in neural modeling (Paninski 2004) and our accompanying Python package supports this  
136 option. To optimize the response template we utilize standard methods—least-squares solvers when minimizing  
137 squared loss, and gradient-based optimization when minimizing the Poisson objective function. To optimize the  
138 warping functions, we found that gradient-based methods were prone to converge to suboptimal local minima, and  
139 that it was preferable and tractable to use randomized searches over these low-dimensional parameter spaces (see  
140 *Methods*).



**Fig 1. Illustration of time warping models.** (A) Synthetic data from a single neuron on 100 trials. When the data are aligned (top left) the trial-average provides an effective, denoised description of activity (bottom left). When the data are misaligned by introducing jitter (right), the trial-average does not capture the typical firing pattern (bottom right). (B) Time warping models estimate a template (black line, left) that can be transformed on a trial-by-trial basis by warping functions (middle column) producing single-trial estimates of neural activity (right column). Red, blue and green lines represent warping functions and estimated firing rates on three example trials. Top row illustrates shifting the template activity in time (shift-only warping), second row illustrates stretching and compressing the template (linear warping), third row illustrates piecewise-linear warping with two line segments (one may increase the number of segments to introduce more nonlinearity into the warping function), the bottom row illustrates a fully nonlinear warping. (C) Results of a shift-only warping model (top), a linear warping model (middle), and a nonlinear warping model (bottom; [Petitjean et al. 2011](#)) fit to the synthetic data from panel A. For each model, we show the warping templates (left, black lines) and warping functions (right, lines colored by ground truth shift to the right vs left). (D) The predicted firing rates of the shift-only warping model provide a denoised trial-by-trial estimate of neural activity. (E) The raw data sorted by the per-trial shift parameter learned by the shift-only warping model. (F) Data aligned by the shift-only model; the per-trial shift learned on the model template is applied in the opposite direction to the raw data.

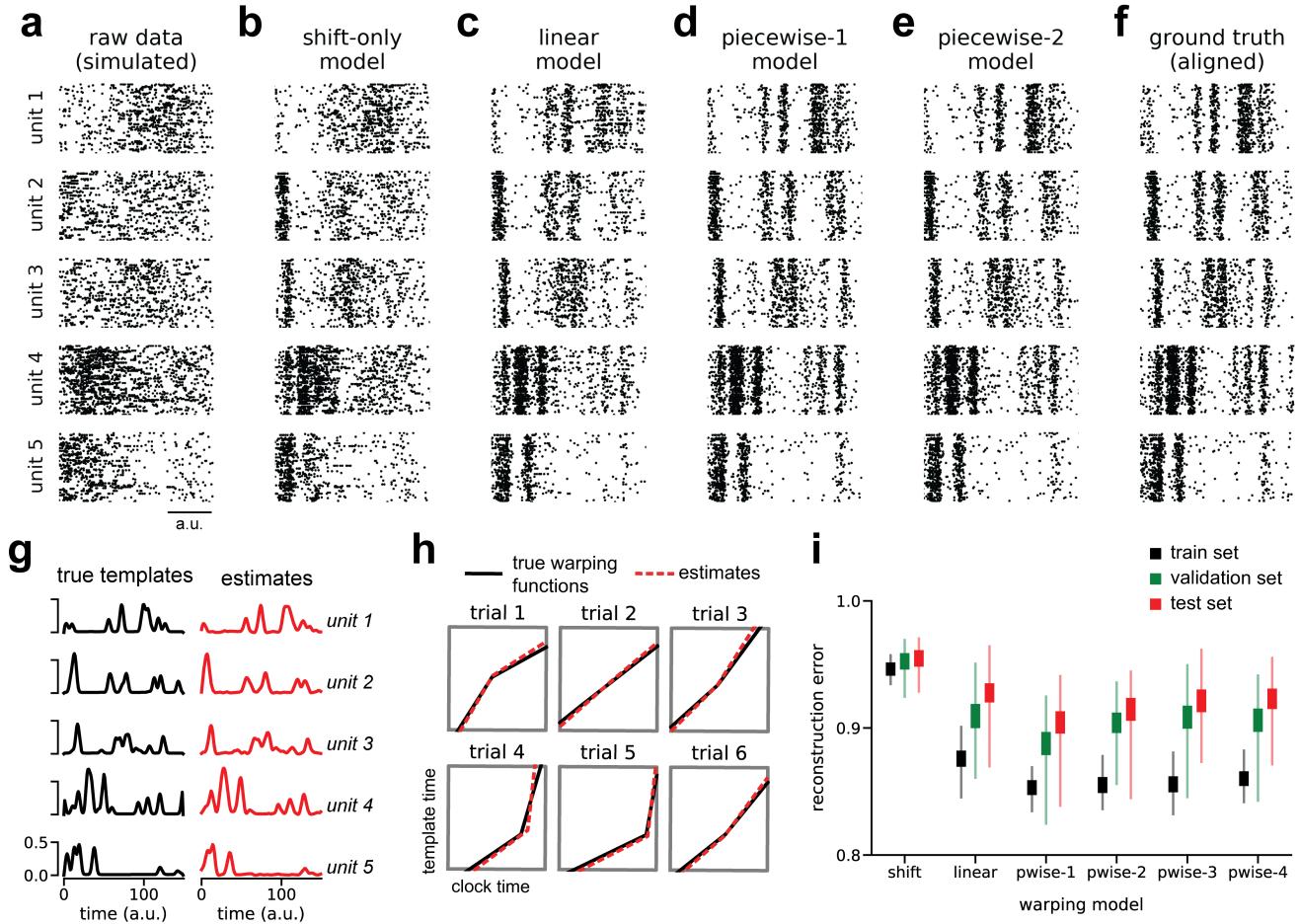
141 For illustration, we fit shift-only, linear, and nonlinear warping models to the misaligned synthetic data shown **Figure**  
142 **1a**. By design, the shift-only model is sufficient to capture the ground truth variability in timing; as expected, this  
143 model identifies a highly accurate template firing pattern (**Fig 1c**, top left), along with warping functions that tightly  
144 correlate with the ground truth delay on a trial-by-trial basis (**Fig 1c**, top right). The linear warping model is a gentle  
145 extension of the shift-only warping model, which only introduces one additional parameter on each trial—the slope  
146 of each warping function. Yet, even this very minor extension produces a slightly worse estimate of the template  
147 and ground truth warping functions (**Fig 1c**, middle). This worsened estimate results from the model overfitting to  
148 noise in the simulated data; the linear warping model can use its additional per-trial parameter to align patterns of  
149 noise in the data across trials, which then appear in the response template. To demonstrate a more severe case  
150 of overfitting, we fit a nonlinear warping model using Dynamic Time Warping (DTW; [Berndt and Clifford 1994](#)),  
151 combined with a standard barycenter averaging procedure ([Petitjean et al. 2011](#)). This method can be highly  
152 effective on datasets with low levels of noise and complex temporal deformations. However, as we will soon see,  
153 neural datasets often exhibit the opposite—high levels of noise and simple temporal deformations. In this regime,  
154 DTW barycenter averaging identifies a noisy and deformed template (**Fig 1c**, bottom left) and warping paths that  
155 correlate with the ground truth jitter, but are unnecessarily complex (**Fig 1c**, bottom right).

156 These results demonstrate that time warping models can be sensitive to noise, especially when a more flexible  
157 class of warping functions is utilized. Our time warping framework uses three strategies to prevent overfitting. First,  
158 as illustrated by the progression of models in **Figure 1c**, we always compare the estimates of complex warping  
159 models (e.g. with piecewise linear warping functions) to the performance of simpler models (e.g. shift-only warping).  
160 Second, we include a smoothness regularization term on the template, which penalizes the average norm of the  
161 second temporal derivative, and thus encourages temporally smooth model estimates. Third, we place a penalty on  
162 the area between each warping function and the identity line, which penalizes the magnitude of warping on each  
163 trial. We include these roughness and warp-magnitude penalties in subsequent results, but show the results of  
164 unregularized time warping in **Figure 1c** for the sake of illustration. A detailed description of these regularization  
165 terms is provided in the *Methods* section.

166 These time warping models enable several strategies for visualizing and understanding neural data. First, one can  
167 directly inspect the model parameters (**Fig 1c**). The response template for each neuron captures the shape of the  
168 neural response, while the warping functions capture trial-to-trial variability. When simple warping functions are  
169 used, the parameters of each function (e.g. the slope and intercept, for a linear warping model) can be visualized in  
170 a histogram or scatterplot, or regressed against behavioral covariates. Second, one can view the model prediction  
171 as a denoised estimate of firing rates on a single-trial basis (**Fig 1d**). Third, one can re-sort the trials by the slope  
172 or the intercept of the warping function, producing a multi-trial raster plot that is easier to visually digest (**Fig 1e**).  
173 Finally, one can invert the learned warping functions on each trial to transform the raw data into an aligned time  
174 domain (**Fig 1f**). This alignment procedure simply entails plotting each activity trace as a function of  $\omega_k^{-1}(t)$  instead  
175 of raw clock time,  $t$ . Intuitively, this amounts to reversing the flow diagram shown **Figure 1b**, which is possible as  
176 long as the warping functions are monotonically increasing (i.e. invertible).

## 177 Extraction of precise, ground truth spike patterns on synthetic data

178 Before proceeding to biological data, we examined a more challenging synthetic dataset involving multiple neurons  
179 and complex single-trial variability in timing. We simulated spike train data from  $N = 5$  neurons,  $T = 150$  timebins,  
180 and  $K = 75$  trials. On each trial, the neural firing rates were time warped by randomized piecewise linear functions  
181 with one knot (the “ground truth” model; see *Methods*). This resulted in spike trains that appear highly variable in  
182 their raw form (**Fig 2a**).



**Fig 2. Recovery of ground truth warping functions in synthetic data.** (A) Synthetic spiking data from  $N = 5$  units,  $T = 150$  timebins, and  $K = 75$  trials. Data were simulated from a ground truth model with piecewise linear warping functions with 1 knot. (B) Data re-aligned by a shift-only warping model. Re-alignment is implemented by applying the inverse warping transformation to the time base of each trial. (C) Data re-aligned by a linear warping model. (D) Data re-aligned by a piecewise-linear (1-knot) warping model. (E) Data re-aligned by a piecewise-linear (2-knots) warping model. (F) Data re-aligned by the ground truth model. Note similarity with panels D & E. (G) Ground truth neural response templates (black) and estimated response templates (red) from the piecewise-linear (1-knot) model. Y-axis denotes the probability of spiking in each time bin. (H) Ground truth warping functions (black) on six representative trials, and estimated warping functions (red) from the piecewise-linear (1-knot) model. (I) Normalized reconstruction errors (norm of residuals divided by norm of raw data) in training, validation, and test partitions of various warping models. Thin lines represent maximum and minimum values; thick lines represent mean  $\pm$  standard error. Results were computed over 40 randomized cross-validation runs.

Given these noisy observations, time warping successfully reveals the spike patterns corresponding to the ground truth process. **Figures 2b-e** show model-aligned spike trains (as in **Fig 1f**), across warping models of increasing complexity. These model-derived alignments can be compared to the ground truth spike times after omitting temporal variability from the simulation (**Fig 2f**). The patterns evident in the ground truth data are partially revealed by shift-only and linear time warping (**Fig 2b-c**), but these models are too simplistic to capture the complete fine-scale temporal structure in the data. A piecewise linear warping model with one knot (*piecewise-1* model; **Fig 2d**) accurately captures these details, and represents a parsimonious and “correct” model since it matches the data generation process. Furthermore, the parameters this model closely matched the ground truth response template (**Fig 2g**) and warping functions (**Fig 2h**). Using a slightly more complex model—a piecewise linear model with 2 knots—did not result in substantial overfitting and indeed closely matched the result of the piecewise-1 model (**Fig 2e**).

Identifying a parsimonious warping model is challenging in real-world applications where there is no observable ground truth. To select the appropriate model and regularization strengths we developed a nested cross-validation scheme (see *Methods*). Briefly, we fit the neural response template using a subset of trials, and we fit the warping functions using a subset of the neurons in the data set (the *training set*). For each warping model class (shift-only, linear, piecewise linear, etc.) we select regularization parameters based on a different subset of neurons and trials (the *validation set*). Finally, we evaluate and compare model performance on the remaining data (the *test set*). This procedure is then repeated many times with different partitions of the data. On simulated data, this procedure correctly identifies the piecewise-1 model as having minimal average test error (**Fig 2i**).

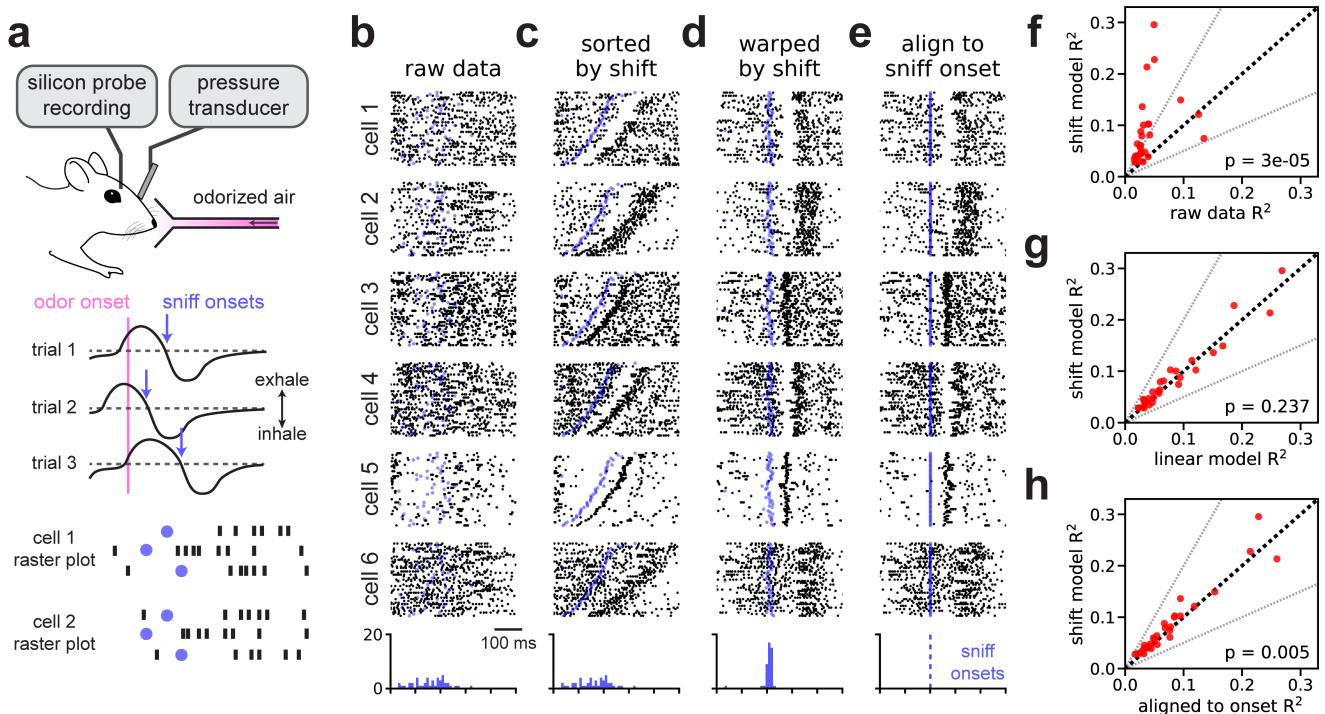
In the following sections, we examine the utility of time warping on real neural datasets derived from a variety of sensory and motor areas. The dynamics of these circuits is thought to be closely time-locked to behaviors and sensory cues, yet we found time warping revealed additional temporal structure and precision in all cases, and even identified unexpected oscillatory patterns in two datasets that were decoupled from measured behaviors. Furthermore, we show that simple time warping models (linear or shift-only) are often sufficient to extract these insights, obviating the need for complex, nonlinear warping methods.

## Alignment of olfactory responses to sniff cycle

Mitral/tufted cells in the mouse olfactory bulb display highly variable firing patterns across trials when naïvely aligned to odor delivery (**Fig 3a**). This variability largely stems from trial-to-trial variability in the latency between odor delivery and inhalation, which controls the access of odorants to receptors. Aligning spike times on each trial to inhalation onset reveals a drastically more reliable encoding of the olfactory stimulus (Shusterman et al. 2011).

We reasoned that simple time warping models could be used to accurately align mitral/tufted cell activity using purely neural activity, bypassing the need to measure inhalation directly. We tested this hypothesis on a multielectrode recording from  $N = 30$  neurons over  $K = 45$  trials of odor presentation at a fixed concentration ( $\alpha$ -pinene,  $10^{-2}$  M). We found comparable results on a separate set of trials on which a different odorant was presented (**Fig 3, Supplement 1**; limonene,  $10^{-2}$  M). We experimentally measured intra-nasal pressure to detect sniff onset and offset. Critically, sniff measurements were not provided to the model and spike times were instead aligned to odor presentation. As expected, this initial alignment strategy produced highly disordered spike rasters for individual neurons (**Fig 3b**).

We found that a shift-only time warping model captured precise sensory responses from these raw data, as revealed by re-sorting the trials based on the model’s shifts (**Fig 3c**) or by applying these shifts to align the raw spike times (**Fig 3d**). Here, as well in all subsequent results, we adopted a leave-one-out validation procedure such that



**Fig 3. Time warping of mitral/tufted cell activity recovers sniff-locked activity patterns.** (A) A head-fixed mouse sampled odorized air ( $\alpha$ -pinene,  $10^{-2}$  M) while spiking activity of isolated mitral/tufted cells neural activity was recorded. Airflow was switched from a non-odorized source to an odorized source on each trial. Variability in inhalation onset from trial-to-trial caused jitter in the onset of the olfactory response. (B) Spike raster plots for six representative cells over all  $K = 45$  trials with spike times aligned to odor delivery. Black dots denote spike times, blue dots denote sniff onset. Blue histogram at the bottom indicates distribution of sniff onset times. (C) Same plots as shown in panel B, but with trials re-ordered by the magnitude of the warping model. (D) Raster plots with spike and sniff onset times re-aligned by applying the inverse warping functions on each trial. (E) Raster plots with spike re-aligned to sniff onset. (F) Trial-to-trial reliability ( $R^2$ ) for all  $N = 30$  cells before (x-axis) and after (y-axis) alignment by the shift-only model. Dashed black line indicates the identity line. Dashed grey lines indicate a two-fold increase or decrease in  $R^2$ . The p-value is computed using a Wilcoxon signed rank test. (G) Same as F, but comparing the alignment of the linear warping model (x-axis) to a shift-only warping model (y-axis). (H) Same as F, but comparing the alignment of the shift-only warping model (y-axis) to the data manually aligned to sniff onset (x-axis).

model-aligned spike rasters were computed only for held out neurons. That is, we excluded each cell (1-5 in **Figure 3**) from model fits, and then simply applied each trial's inverse warping function to the held out cell. Thus any temporal structure seen in **Figure 3d** is unlikely to arise as an artifact of overfitting.

As expected, aligning spike times to inhalation onset times reveals similar patterns in these data (**Fig 3e**). Indeed, the shift parameters learned by the model correlated very tightly with the onset of sniffing (see blue dots and histograms, **Fig 3b-e**). These results are nonetheless a useful demonstration, since the model inferred these precise responses from the neural data alone and without reference to intra-nasal pressure. Furthermore, closer examination suggested that the unsupervised, shift-only model may enjoy slight performance advantages relative to the simple supervised alignment method. For example, when aligned to sniff onset, cells 4 and 5 in **Figure 3** exhibit subtle, but perceptible, jitter in their responses (**Fig 3e**), and this variability is visibly corrected by time warping (compare to **Fig 3d**).

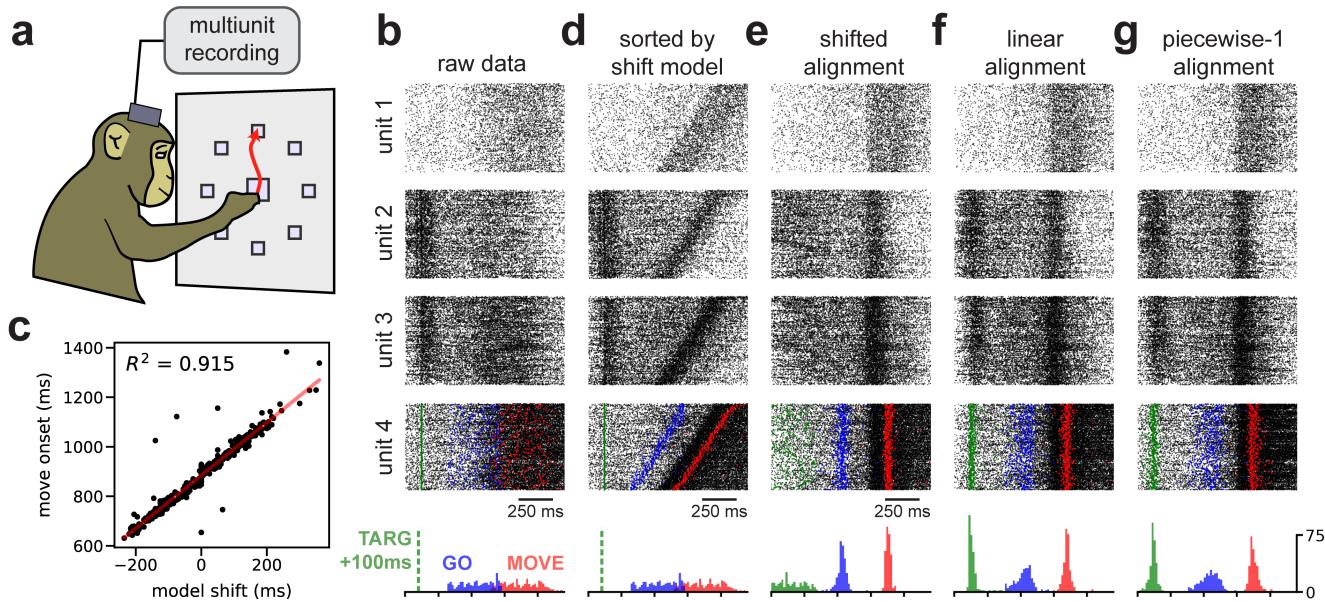
We quantified the trial-to-trial variability of each neuron by computing the coefficient of determination ( $R^2$ ) of the neuron's PSTH. In an approach similar to leave-one-out cross-validation, we fit time warping models while holding out neurons one at a time; the  $R^2$  was then computed on the held out neuron before and after warping. Relative to the raw spike times (i.e. odor onset aligned), shift-only time warping improved  $R^2$  in nearly all neurons, with many increasing over two-fold (**Fig 3f**; average 107% increase in  $R^2$ , geometric mean; Wilcoxon signed rank test,  $p < 10^{-4}$ ,  $n = 30$ ). Moreover, moving from a shift-only warping model to a more flexible linear warping model did not produce any significant improvements in  $R^2$  (**Fig 3g**). Relative to sniff onset alignment, shift-only time warping improved the  $R^2$  criterion mildly (**Fig 3h**; average 11% increase in  $R^2$ , geometric mean; Wilcoxon signed rank test,  $p = 0.005$ ,  $n = 30$ ).

## Alignment of motor cortex dynamics during reaching in nonhuman primates

Neural dynamics underlying motor control also exhibit variable time courses due to trial-to-trial differences in reaction times and muscle kinematics. To investigate the benefits of time warping in this setting, we first examined data from a canonical reaching experiment in a nonhuman primate (**Fig 4a**). On each trial, the subject (Monkey J) moved its arm to one of several possible target locations after a mandatory delay period that randomly varied between 300 and 700 ms. In addition to this inherent timing variability due to task design, the monkey exhibited variable reaction time ranging from 293-442 ms (5th and 95 percentiles). We limited our analysis to upward reaches (90° from center) with the target placed at 40, 80, or 120 cm from the center. We observed similar results on other reach angles (**Fig 4, Supplement 1**), as well as when data was pooled across all reach angles (data not shown). Multiunit activity was collected from  $N = 191$  electrodes across two Utah multielectrode arrays placed in primary motor (M1) and premotor (PMd) cortices (see *Methods*).

The most dramatic changes in neural firing rates are closely time-locked to movement (Churchland et al. 2012; Kaufman et al. 2016). Thus, it is common to track hand position on a moment-by-moment basis and use these measurements to align spike times to the onset of movement or the peak hand velocity on each trial. We instead examined spike trains aligned to the beginning of the delay period (**Fig 4b**), and used time warping to infer an alignment without any reference to the animal's behavior.

As expected, a shift-only warping model closely aligned spike times with the onset of movement. The model's learned shift parameter on each trial correlated very tightly with movement onset (**Fig 4c**), achieving a comparable level of performance ( $R^2 = 0.9$ ) to what was recently reported for a complex, nonlinear warping method (Duncker and Sahani 2018). Furthermore, the shift-only warping model enabled the visualization of movement-related firing rate changes in single-neuron rasters, either by re-sorting the trial order of the raw data (**Fig 4d**) or by re-aligning



**Fig 4. Time warping of reach dynamics in a nonhuman primate.** (A) Reaches towards a 90° target were analyzed. (B) Spike data from four example multiunits over all trials. Units 1 & 4 were from primary motor cortex; units 2 & 3 were from pre-motor cortex. At the bottom, the time distribution of task events is shown: 100 ms after target onset (TARG, green), end of delay period (GO, blue), and movement onset (MOVE, red). (C) Scatterplot showing correlation between the hand movement onset on each trial and the learned shift parameter in the shift-only warping model. (D) Same as panel B, but with trials re-sorted by the per-trial shift parameter in the shift-only warping model. (E) Same as panel B, but with spikes aligned according to the shift-only warping model. (F) Same as panel B, but with spikes aligned according to the linear warping model. (G) Same as panel B, but with spikes aligned according to the piecewise-linear warping model with one knot.

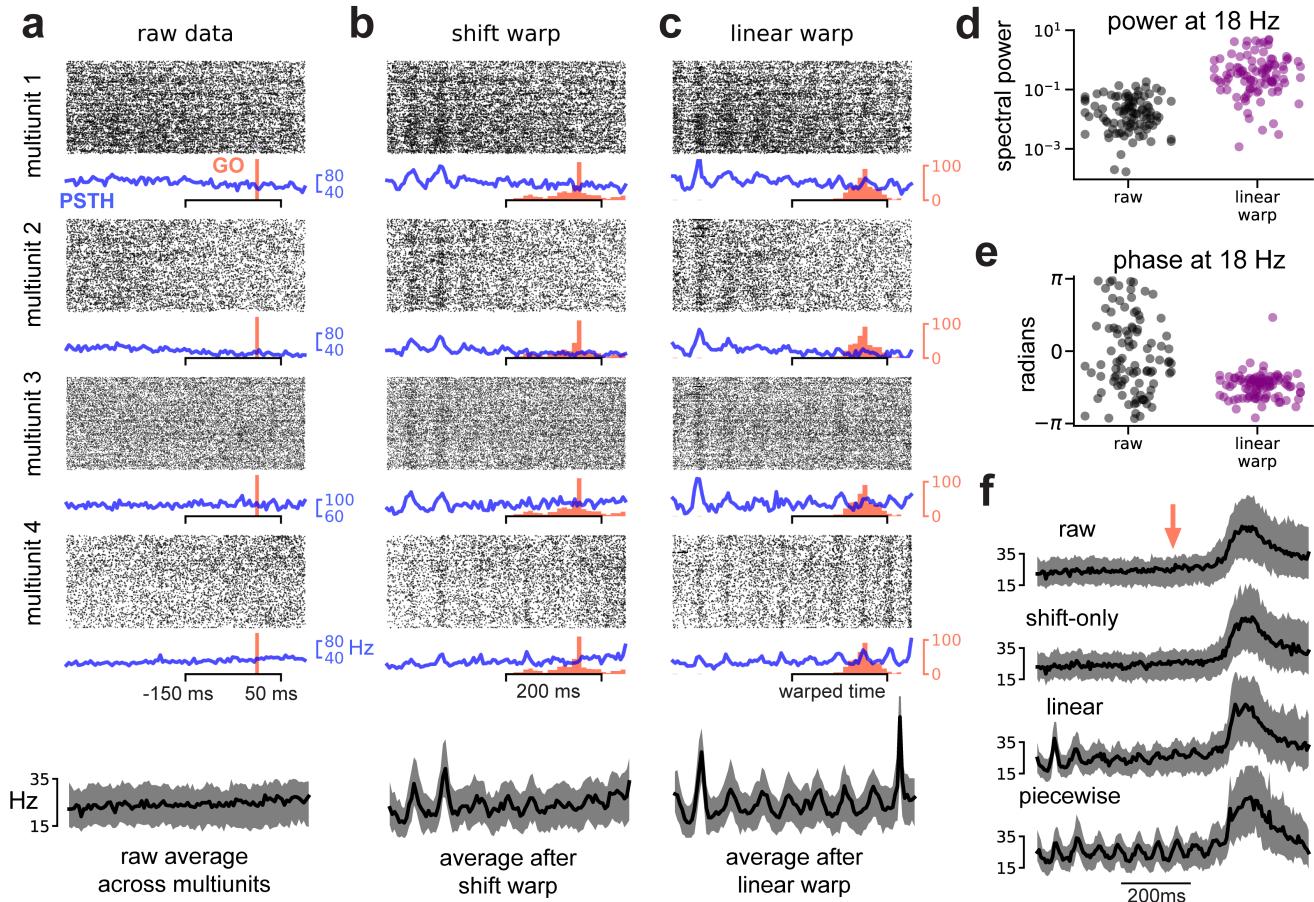
265 the spike times (Fig 4e).

266 Thus, learning a single per-trial shift was sufficient to align neural spike times to movement, without any reference to  
267 hand tracking data. However, shifting spike times in this manner also *destroyed* other structure in the data. Namely,  
268 a subset of multiunits, mostly in PMd, showed increased firing around ~100ms into the delay period—i.e. shortly  
269 after the reach target was visually presented to the animal (see units 2 & 3 in Fig 4). Due to the variable delay  
270 between target onset and movement onset, a shift-only warping model is incapable of simultaneously aligning  
271 spikes across these two events.

272 A linear time warping model more appropriately captures this structure in the data. On each trial, the model utilizes  
273 its two free parameters—the slope and intercept of the warping function—to precisely align these two task events  
274 (Fig 4f). Importantly, as in all of our results, the warping model is fit purely to the neural data without any reference to  
275 behavior. Thus, these results provide strong evidence, via an unsupervised time warping method, that reliable neural  
276 dynamics occur around the time of movement onset and shortly after target onset. Using nested cross-validation,  
277 we determined that more complex, piecewise-linear warping functions did not provide large benefits over the linear  
278 warping model; however, the linear warping model provided a reproducible benefit over the shift-only warping model  
279 (Fig 4, Supplement 2).

## 280 Detection of ~13-40 Hz spike-time oscillations in primate pre-motor cortex

281 Thus far, we have shown that the temporal alignments learned by simple warping models can closely correlate  
282 with behaviors (e.g. movement or sniffing) and sensory cues (e.g. reach target presentation). This agreement



**Fig 5. Spike-level oscillations in primate premotor cortex.** (A) Multiunit activity aligned to go cue. Four representative multiunits are shown; spike rasters (black) trial-average PSTH (blue) are shown for each multiunit. Go cue onset is shown in red. The average firing rate across all multiunits is shown on the bottom (black trace). Shaded grey region denotes upper and lower quartiles. (B) Same as (A) but with spike times aligned by shift-only warping model. Each displayed multiunit was held out from the model fit. (C) Same as (A) but with spike times aligned by linear warping model. Each displayed multiunit was held out from the model fit. (D) Oscillatory power at 18Hz in trial-average multiunit activity aligned to go cue (black dots) and after linear warping (purple dots). Each dot represents one of  $N = 96$  multiunits. (E) Same as (E) but showing oscillation phase for each multiunit. (F) Multiunit activity averaged across electrodes and trials in a larger time window (800 ms) around go cue (red arrow). Oscillations are not visible in the raw data (top) or after shift warping (second from top). Oscillations are recovered by either linear (second from bottom) or piecewise linear warping (bottom). Shaded grey region denotes upper and lower quartiles.

283 demonstrates that time warping models can converge to reasonable and human-interpretable solutions, and,  
284 conversely, suggests that established alignment practices in these systems are well-justified from a statistical  
285 perspective. However, time warping methods can also uncover more subtle and unexpected features in spike train  
286 data.

287 In primate premotor cortex, the local field potential (LFP) shows prominent oscillations in the beta frequency range  
288 (13-40 Hz) during movement preparation, which are correlated with spike timing (Murthy and Fetz 1992; Sanes and  
289 Donoghue 1993; Reimer and Hatsopoulos 2010). While recent work has elucidated the statistical relationships  
290 between LFP and behavior (Khanna and Carmena 2015; Khanna and Carmena 2017; Chandrasekaran et al.  
291 2019), the impact of beta oscillations on population-level spiking activity is still poorly understood. Recent work  
292 used a complex, black box model of neural dynamics to detect oscillatory structure in high-dimensional spike  
293 trains (Pandarinath et al. 2018). Here, we show that shift-only or linear time warping models can recover similar  
294 oscillations, and compactly summarize single-trial variations in oscillation phase and frequency in the warping  
295 function parameters.

296 We examined premotor cortical data collected from two different monkey subjects (Monkey J and Monkey U)  
297 performing point-to-point reaches; one animal performed these reaches with an unrestrained hand, while the other  
298 used a manipulandum (see *Methods*). The oscillations are strongest during the pre-movement delay period, and  
299 thus we first focused on a time window beginning 400 ms prior to and 100 ms after go cue presentation. We found  
300 that having a larger number of trials was beneficial, so we pooled trials from all reach angles for this analysis. We  
301 analyzed multiunit data (not spike sorted) for each monkey from  $N = 96$  electrodes placed in PMd.

302 No oscillations were visible in pre-movement spike rasters aligned to go cue (**Fig 5a**; data from Monkey U).  
303 However, re-aligning these spike trains based on a shift-only warping model revealed oscillations in virtually all  
304 multiunits. These oscillations occurred at  $\sim 18$  Hz in Monkey U (shown in **Fig 5b**) and at  $\sim 40$  Hz in Monkey J (**Fig 5**,  
305 **Supplement 1**); these results are within previously reported frequency ranges (Murthy and Fetz 1992; Sanes and  
306 Donoghue 1993). In Monkey U, these oscillations were more apparent after linear warping (**Fig 5c**), suggesting that  
307 the frequency (in addition to the phase) of the oscillations can be variable on each trial. These oscillations were  
308 roughly in-phase across multiunits—as a result, averaging spike counts across all multiunits and trials (**Fig 5a-c**,  
309 bottom) produced a cleaner  $\sim 18$  Hz oscillation in time warped spike trains.

310 We confirmed that the spike-level oscillations were in-phase with LFP oscillations in Monkey U. To do this, we  
311 applied the time warping models fit on spike train data to bandpass-filtered LFP signals (10-30 Hz). The LFP signal  
312 was misaligned across trials in raw data, but was accurately aligned by the spike-level time warping models (**Fig**  
313 **5, Supplement 2**), suggesting that the two signals are coherently time-warped (in this case, temporally shifted  
314 and/or stretched) on a trial-by-trial basis. On a methodological level, this demonstrates that time warping models  
315 can *generalize* and make accurate predictions about other time series (e.g., LFP) with qualitatively distinct statistics  
316 from the training data (e.g., spike times). This ability to identify structure across different data streams in a flexible,  
317 unsupervised manner is an attractive feature of time warping models, which is facilitated by our choice to use simple  
318 and invertible warping functions.

319 To quantify these effects more carefully across all multiunits, we compared the PSTHs computed from raw data  
320 (blue traces; **Fig 5a**) to PSTHs computed from data aligned by linear time warping (blue traces; **Fig 5c**). The  
321 raw PSTHs exhibited no oscillations, as this pattern was temporally jittered and stretched from trial-to-trial and  
322 therefore abolished by trial averaging. In contrast, oscillations can be observed to varying extents in the PSTHs  
323 computed after alignment by linear warping (which corrects for these trial-to-trial variations). Using Fourier analysis  
324 to estimate the amplitude and phase of the oscillation at 18 Hz, we found that alignment by linear warping increased  
325 the strength of the oscillation by 1-2 orders of magnitude in most multiunits (**Fig 5d**). Furthermore, in the raw PSTHs  
326 the oscillation phases were widely spread across multiunits, consistent with there being no detectable oscillations

327 above background noise (**Fig 5e**; gray dots); in the aligned PSTHs, the phases were tightly clustered, reflecting that  
328 nearly all multiunits oscillated in a coherent and detectable manner (**Fig 5e**; purple dots).

329 We wondered whether time warping would fail to recover these oscillations if the movement-related spiking, which  
330 occurs at a much higher firing rate than pre-movement activity, was included in the analysis. To examine this, we fit  
331 warping models to a larger time window ( $\pm 400$  ms around go cue presentation), which included the movement-  
332 related increase in firing rate. Time warping was still able to extract oscillations under these more challenging  
333 circumstances (**Fig 5f**). Interestingly, a shift-only model was no longer sufficient to reliably capture oscillatory activity,  
334 suggesting that the oscillations were not phase-aligned with movement onset on a trial-by-trial basis. In contrast,  
335 linear or piecewise linear warping functions were able to recover the oscillatory structure (**Fig 5f**; bottom). Thus,  
336 while the shift-only model is simplest to interpret, it may be insufficient to capture certain results under particular  
337 circumstances. This emphasizes the utility of conceptualizing time warping as a range of models (as in **Figure**  
338 **1b**) rather than a single method—one can systematically increase the warping complexity to capture increasingly  
339 complex features in neural data.

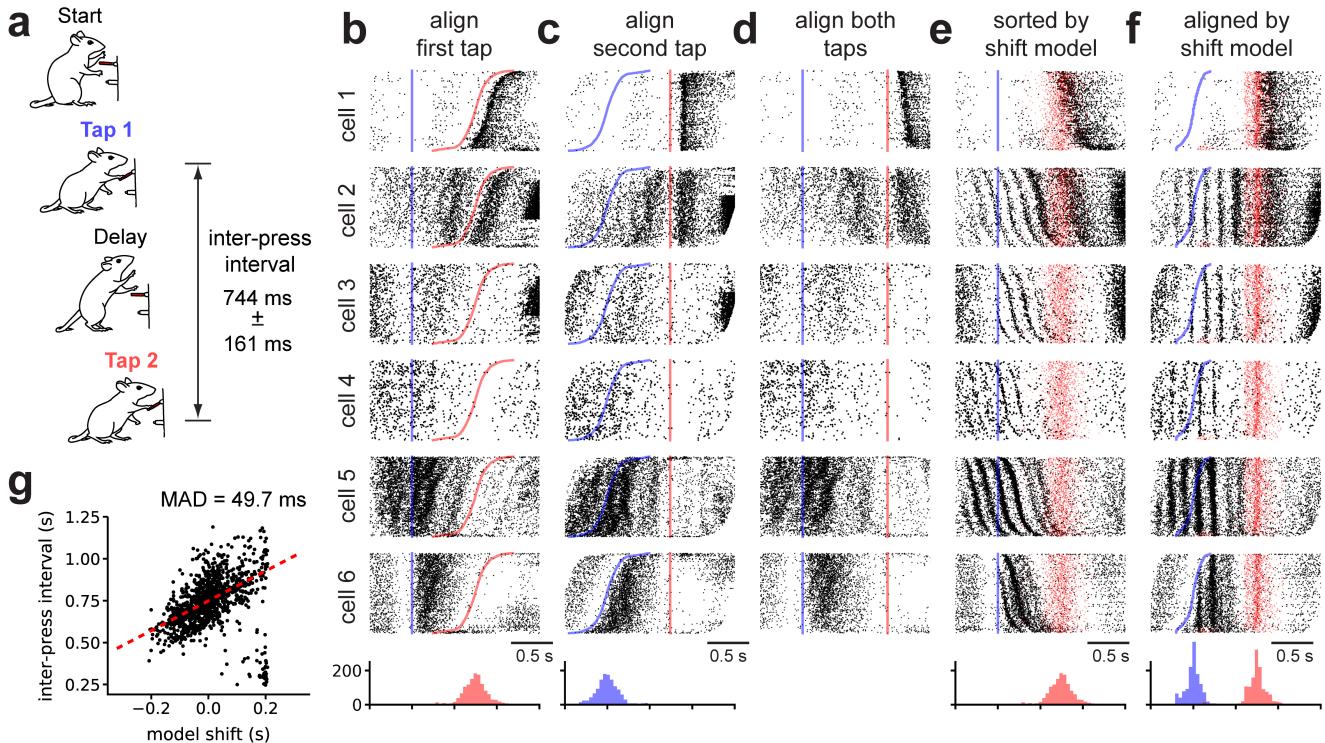
### 340 Detection of ~6-7 Hz oscillations in rat motor cortex

341 We have seen that time warping can reveal interpretable structure, even under very simple and well-controlled  
342 experimental conditions. Discrete reaching, for example, is arguably the simplest volitional motor behavior that  
343 one can study, and yet straightforward behavioral alignments obscure salient spike-time oscillations (see **Fig 5**).  
344 To study a more complex behavior, in a different animal model, we analyzed motor cortical activity in rats trained  
345 to produce a timed motor sequence (Kawai et al. 2015; Dhawale et al. 2017). Rats were trained to press a lever  
346 twice with a target time interval of 780 ms, and were rewarded if the sequence was completed within  $\pm 80$  ms of  
347 this target (**Fig 6a**). While rats produce stereotyped motor sequences in this setting, the duration between lever  
348 presses and the timing of intermediate motor actions is variable from trial-to-trial. We examined a dataset consisting  
349 of  $N = 30$  neurons and  $K = 1265$  trials; the interval between lever presses ranged from 521-976 ms (5th- and  
350 95th-percentiles) across trials.

351 This experiment has three obvious alignment procedures: align spike times to the first lever press, align spike times  
352 to the second lever press, or linearly stretch/compress the spike times to align both lever presses across trials (i.e.  
353 human-supervised time warping). **Figure 6b-d** shows the activity of six example neurons under these alignment  
354 strategies. At a high level, these rasters demonstrate that neurons preferentially respond to different behavioral  
355 events within a trial. For example, cell 1 in **Figure 6** fires after the second lever press, while cell 6 in **Figure 6** fires  
356 after the first press. Thus, it is not obvious which alignment is preferable and indeed different insights may be gained  
357 from analyzing each.

358 Unsupervised time warping revealed structure in the data that is hidden in all three behavioral alignments. A  
359 shift-only warping model uncovered strong oscillations in many neurons, as visualized either by re-sorting trials  
360 based on the learned shift (**Fig 6e**, same alignment as **Fig 6b**), or by using the model to re-align spike times (**Fig**  
361 **6f**). These findings are not due to spurious alignments produced by an overfit model. Each spike raster in **Figure 6f**  
362 was generated from data held-out from the model—that is, the visualized cells were excluded during optimization of  
363 the warping functions, and the learned alignment transformation was then applied to this held out data.

364 These results reveal a partial decoupling of behavioral events (lever presses) with neural firing patterns. After  
365 alignment, both the first and second lever presses occur at variable times within each trial (**Fig 6f**, histograms at  
366 bottom). Furthermore, the learned shift on each trial only loosely correlated with inter-press interval (**Fig 6g**). Taken  
367 together, these features of the data suggest that it would be difficult to discover this oscillatory structure by manual



**Fig 6. Shift-only time warping reveals temporally precise theta-locked oscillations in rat motor cortex.** (A) Rats were trained to press a lever twice with a prescribed temporal delay. The median  $\pm$  IQR inter-press-interval is listed. (B) Spike raster plots for six representative cells over all trials with spike times aligned to the first lever press (blue line) and trials sorted by the inter-press-interval. The time of the second lever press is denoted by the red line in all plots. Red histogram at the bottom denotes the distribution of the second lever press times. Blue histogram at the bottom denotes the distribution of the first lever press times. (C) Spike raster plots re-aligned to the second lever press. Blue histogram at the bottom denotes the distribution of the first lever press times. (D) Spike raster plots aligned by linearly stretching/compressing the time axis in each trial so that the first and second lever presses were both aligned (note the lack of scale bar, as time is no longer constant across trials). (E) Same spike rasters as in panel B, but with trials re-sorted by the magnitude of a shift-only warping model. The time of the second lever press on each trial is denoted by a semi-transparent red dot. (F) Spike raster plots after re-sorting trials as in panel E and shifting the spike trains on each by the shift-only time warping model. (G) Relationship between per-trial shift learned by the shift-only time warping model (horizontal axis) and measured inter-press interval (vertical axis) on each trial. No tight correlation with behavior is observed—dashed red line denotes a robust linear regression fit (Huber loss function,  $\epsilon = 1.001$ ), and the Median Absolute Deviation (MAD) is listed as a measure of fit.

368 alignment, demonstrating the power of unsupervised time warping models.

369 While the uncovered oscillations are not phase locked with lever press times, they are nonetheless correlated with  
370 certain aspects of the animal's behavior. In particular, some cells only exhibit the ~6-7 Hz oscillation following the  
371 first lever press, with remarkable temporal precision (see cells 3 and 6 in **Fig 6f**). Indeed, multiple cells exhibit  
372 non-oscillatory firing prior to the first lever press, but rapidly switch to an oscillatory behavior following the lever  
373 press (see cell 3 in **Fig 6f** and **Fig 6, Supplement 1**). Other cells exhibit oscillations prior to the first lever press, but  
374 the amplitude and precision of the oscillations appears to improve following the first lever press (see cell 4 in **Fig**  
375 **5F**). Still other cells either do not exhibit oscillations (cell 1 in **Fig 6f**) or exhibit strong oscillations both prior to and  
376 following the first lever press (cell 5 in **Fig 6f**). Time warping enables us to discover and visualize this full spectrum  
377 of functional cell types, which are otherwise difficult to detect and characterize. The presence of oscillations in single  
378 neurons can be confirmed by plotting the distribution of inter-spike-intervals (**Fig 6, Supplement 1**); the shift-only  
379 model goes beyond this method by demonstrating that a large population of neurons are coherently phase-shifted  
380 on a trial-by-trial basis, and by enabling characterization of the full population dynamics and behavioral events in an  
381 aligned temporal space.

382 We then examined whether these spike-level oscillations were aligned with oscillations in LFP. The average frequency  
383 spectrum of the LFP did display a prominent peak at ~6-7 Hz—a very similar frequency range to the spike-level  
384 oscillations identified in **Figure 6**. To characterize the relationship between these two oscillatory signals, we  
385 bandpassed filtered the LFP between 5-9 Hz on each trial and fit a separate shift-only time warping model to the  
386 LFP traces. The time warping functions learned on LFP data did not uncover the spike-time oscillations shown in  
387 **Figure 6**, and likewise the LFP signals were not well-aligned by the time warping functions fit to spike times (**Fig**  
388 **6, Supplement 1**). Thus, unlike the oscillations identified in primate premotor cortex, the oscillations in rat motor  
389 cortex were not aligned with the LFP. This analysis illustrates two useful features of our time warping framework.  
390 First, the models can be flexibly applied to other data types beyond spiking data (e.g. LFP). Second, if two or  
391 more data types are simultaneously collected (e.g., LFP and spikes), separate time warping models can be fit to  
392 each signal and then compared to assess whether these signals are coherently warped on a trial-by-trial basis.  
393 These post-hoc comparisons are drastically simplified when time warping functions are constrained to be linear or  
394 piecewise linear, instead of fully nonlinear.

## 395 Discussion

396 While the temporal precision of neural coding has been a matter of intense debate, few studies have leveraged  
397 statistical alignment methods to address this problem. Earlier work incorporated time warping into single neuron  
398 encoding and decoding models (Aldworth et al. 2005; Gollisch 2006; Smith and Paninski 2013; Lawlor et al. 2018),  
399 as well as dimensionality reduction methods (Poole et al. 2017; Duncker and Sahani 2018). Here, we decoupled  
400 time warping from these other modeling objectives, to achieve a flexible and simplified framework. We surveyed a  
401 broader range of datasets than past work, spanning multiple model organisms, brain areas, and sensory/motor  
402 tasks. In all cases, we found that the simplest and most interpretable models—often those with shift-only or linear  
403 warping functions—matched the performance of more complex models, while uncovering striking and sometimes  
404 unanticipated dynamics.

405 We first examined two datasets in which behavioral alignments are well-established, and found that unsupervised  
406 time warping inferred similar alignments based on neural data alone. In mouse olfaction, mitral/tufted cells exhibit  
407 reliable sensory coding when spike times are aligned to sniff onset, but not odor onset (Shusterman et al. 2011). The  
408 simplest time warping model—a shift-only model—accurately realigned spikes without any reference to intranasal

409 pressure measurements, and more complex warping methods (linear and piecewise linear) produced little additional  
410 benefit. In primates executing cued reaches, a shift-only time warping model accurately predicted movement  
411 onsets ( $R^2 \approx 0.9$ ) without any reference to hand tracking data. This performance is comparable to recent work that  
412 developed a nonlinear warping method (Duncker and Sahani 2018), and surpasses prior work (Petreska et al. 2011;  
413 Poole et al. 2017).

414 Together, these results demonstrate that shift-only and linear warping models can match or even outperform more  
415 complex methods. These simpler models have two attractive properties. First, they manipulate model estimates  
416 of single-trial firing rates in a more interpretable manner (see Fig 1), enabling exploratory data analysis and  
417 visualization. Second, we developed fast and computationally scalable optimization methods for this class of models.  
418 On a modern laptop, these models can typically be fit to data from 1000 neurons, 100 timepoints, and 1000 trials  
419 in one minute or less. This scalability is of great practical importance given the exponentially increasing size of  
420 neural recordings (Stevenson and Kording 2011), and the growing need for rigorous cross-validation and model  
421 comparison methods (Chandrasekaran et al. 2018), which are often computationally intensive if not prohibitive.

422 Time warping also uncovered firing patterns that were not aligned to any stimulus or measured behavior. For  
423 example, we observed  $\sim$ 13-40Hz spike time oscillations in primate premotor cortex during movement preparation  
424 (see Fig 5), which we then verified were phase-aligned with LFP (see Fig 5, Supplement 2). Notably, the time  
425 warping models we used did not assume any oscillatory structure in the data, and thus provide a data-driven  
426 validation that spike-level oscillations are a salient feature of the dynamics. Furthermore, a linear time warping  
427 considers the activity of the full neural population to estimate the changes in the phase (y-intercept) and frequency  
428 (slope) of the oscillation on a trial-by-trial basis. This population-level approach can be contrasted with popular  
429 frequency-domain statistical measures like *coherence*, which measures the degree of phase synchronization  
430 between two spike trains or between a single spike train and LFP (Fries et al. 1997; Jarvis and Mitra 2001; Sun  
431 et al. 2005; Aoi et al. 2015). Future work could incorporate oscillatory basis functions into time warping models  
432 to combine the benefits of pairwise spectral analysis with the population-level modeling perspective adopted in  
433 this paper. By drawing statistical power from larger numbers of simultaneously recorded neurons, research in this  
434 direction could provide tighter links between spike-based and LFP-based measures of oscillation, which have been  
435 difficult to characterize despite extensive prior work (Ray 2015).

436 However, oscillatory patterns may not always be synchronized to LFP or pre-conceived behavioral variables, as we  
437 observed in rat motor cortex (see Fig 6). While further work is needed to fully elucidate the properties and functions  
438 of these  $\sim$ 7 Hz oscillations, we found that they were, in some neurons, gated by a motor action—specifically,  
439 the first lever press—suggesting a potential relevance of these oscillations to the motor time keeping task (Fig 6,  
440 Supplement 2). Another, possibility is that orofacial behaviors such as whisking and licking are the primary driver  
441 of these oscillations (Hill et al. 2011). Other work has shown that persistent  $\sim$ 7 Hz LFP oscillations may be locked to  
442 the respiration cycle (Tort et al. 2018); the transient, spike-level oscillations we observed were decoupled from LFP,  
443 and thus likely distinct from this phenomenon. Regardless of their root cause, this result demonstrates the ability of  
444 time warping to extract unexpected features of scientific interest from high-dimensional spike trains. Thus, while it  
445 will be interesting to develop specialized extensions to time warping that address particular scientific questions (e.g.  
446 oscillatory firing patterns), the general-purpose framework developed here can be a powerful tool for exploratory  
447 analysis, as it makes few pre-conceived assumptions about the data.

448 It is possible that future work using more complex, nonlinear warping methods can uncover even finer structure in  
449 neural data. However, we observed that DTW and other classical methods were prone to overfit data, suggesting that  
450 careful regularization will be needed for this approach to succeed. A recently proposed method, called *soft-DTW*,  
451 looks promising (Cuturi and Blondel 2017). While the method is mathematically elegant, we found that soft-DTW  
452 can be difficult to interpret as it does not represent temporal alignments as a single warping function, but rather uses

453 a weighted combination of all possible warping paths. In general, nonlinear warping methods will require careful  
454 application, cross-validation, and secondary analyses to be useful statistical tools for neuroscience.

455 Time warping is only one form of variability exhibited by single-trial neural dynamics. We purposefully examined  
456 time warping models in the absence of other modeling assumptions, such as trial-to-trial variation in amplitude  
457 (Bollimunta et al. 2007; Goris et al. 2014; Williams et al. 2018), or condition-specific changes in dynamics (Duncker  
458 and Sahani 2018). We also made the restrictive assumption that all neurons share the same time warping function  
459 on an individual trial (Shokoohi-Yekta et al. 2015). Finally, we assumed and exploited a trial structure to neural  
460 time series data throughout this work. To study more unstructured time series, future work could incorporate time  
461 warping into state space models (Macke et al. 2015) or sequence extraction algorithms (Mackevicius et al. 2019).  
462 Despite these exciting prospects for future statistical methodology, our work demonstrates that even a simple time  
463 warping framework can provide a rich and practical set of tools for the modern neuroscientist.

464 While our results already show that averaging over short, stereotyped trials can obscure fine temporal oscillations  
465 and firing events, these shortcomings are undoubtedly more severe in behaviors that have longer temporal extents  
466 and exhibit more variability. Thus, we expect time warping methods to play an increasingly crucial role in neural data  
467 analysis as the field moves to study more complex and unstructured animal behaviors (e.g. under more naturalistic  
468 settings; Krakauer et al. 2017). Furthermore, in complex experimental tasks involving large numbers of conditions  
469 and exploratory behaviors, the same motor act or sensory percept may present itself only a small number of times.  
470 In this trial-limited regime, precise data alignment may be critical to achieve the necessary statistical power to make  
471 scientific claims. We expect simple models, such as linear and piecewise linear warping, to perform best on these  
472 emerging datasets due to their interpretability, computational efficiency, and robustness to overfitting.

## 473 Acknowledgements

474 We thank Isabel Low and Chandramouli Chandrasekaran for feedback and detailed discussions on the manuscript.  
475 This work received support from the following research grants and foundations: Department of Energy CSGF  
476 fellowship (A.H.W.), NIH NRSA 1F31NS089376-01 (E.M.T.), Stanford Graduate Fellowship (E.M.T.), NSF IGERT  
477 0734683 (E.M.T.), NIH NINDS T-R01NS076460 (K.V.S.), NIH NIMH T-R01MH09964703 (K.V.S.), NIH Director's  
478 Pioneer Award 8DP1HD075623 (K.V.S.), DARPA BTO "REPAIR" N66001-10-C-2010 and "NeuroFAST" W911NF-14-  
479 2-0013 (K.V.S.), the Simons Foundation Collaboration on the Global Brain awards 325380 and 543045 (K.V.S.),  
480 ONR N000141812158 (K.V.S.), the Howard Hughes Medical Institute (K.V.S.), Charles A. King Trust (A.K.D.), the  
481 Life Sciences Research Foundation (A.K.D.), NIH NINDS R01-NS099323-01 (B.P.O.), NIH R01-DC013797 (D.R.),  
482 NIH R01-DC014366 (D.R.), Burroughs Wellcome (S.G.), Alfred P. Sloan Foundation (S.G.), Simons Foundation  
483 (S.G.), McKnight Foundation (S.G.), James S. McDonell Foundation (S.G.), and the Office of Naval Research (S.G.).

## 484 References

- 485 Aldworth, Zane N., John P. Miller, Tomás Gedeon, Graham I. Cummins, and Alexander G. Dimitrov (2005). "Dejittered  
486 Spike-Conditioned Stimulus Waveforms Yield Improved Estimates of Neuronal Feature Selectivity and Spike-  
487 Timing Precision of Sensory Interneurons". *Journal of Neuroscience* 25.22, pp. 5323–5332.  
488 Amarasingham, Asuhan, Stuart Geman, and Matthew T. Harrison (2015). "Ambiguity and nonidentifiability in the  
489 statistical analysis of neural codes". *Proceedings of the National Academy of Sciences* 112.20, pp. 6455–6460.

- 490 Amarasingham, Asohan, Matthew T. Harrison, Nicholas G. Hatsopoulos, and Stuart Geman (2012). "Conditional  
491 modeling and the jitter method of spike resampling". *Journal of Neurophysiology* 107.2, pp. 517–531.
- 492 Ames, K. Cora, Stephen I. Ryu, and Krishna V. Shenoy (2014). "Neural Dynamics of Reaching following Incorrect or  
493 Absent Motor Preparation". *Neuron* 81.2, pp. 438–451.
- 494 Aoi, Mikio C., Kyle Q. Lepage, Mark A. Kramer, and Uri T. Eden (2015). "Rate-adjusted spike-LFP coherence  
495 comparisons from spike-train statistics". *Journal of Neuroscience Methods* 240, pp. 141–153.
- 496 Arneodo, Ezequiel M, Kristina B Penikis, Neil Rabinowitz, Angela Licata, Annika Cichy, Jingji Zhang, Thomas Bozza,  
497 and Dmitry Rinberg (2018). "Stimulus dependent diversity and stereotypy in the output of an olfactory functional  
498 unit". *Nature Communications* 9.1, p. 1347.
- 499 Aronov, Dmitriy, Rhino Nevers, and David W Tank (2017). "Mapping of a non-spatial dimension by the hippocam-  
500 pal–entorhinal circuit". *Nature* 543, p. 719.
- 501 Baker, Stuart N and George L Gerstein (2001). "Determination of response latency and its application to normaliza-  
502 tion of cross-correlation measures". *Neural Computation* 13.6, pp. 1351–1377.
- 503 Berndt, Donald J. and James Clifford (1994). "Using Dynamic Time Warping to Find Patterns in Time Series".  
504 *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*. AAAIWS'94. Seattle,  
505 WA: AAAI Press, pp. 359–370.
- 506 Bollimunta, Anil, Kevin H Knuth, and Mingzhou Ding (2007). "Trial-by-trial estimation of amplitude and latency  
507 variability in neuronal spike trains". *Journal of neuroscience methods* 160.1, pp. 163–170.
- 508 Brette, Romain (2015). "Philosophy of the Spike: Rate-Based vs. Spike-Based Theories of the Brain". *Frontiers in  
509 Systems Neuroscience* 9, p. 151.
- 510 Bro, R., K. Kjeldahl, A. K. Smilde, and H. A. L. Kiers (2008). "Cross-validation of component models: A critical look  
511 at current methods". *Analytical and Bioanalytical Chemistry* 390.5, pp. 1241–1251.
- 512 Bruno, Randy M (2011). "Synchrony in sensation". *Current Opinion in Neurobiology* 21.5. Networks, circuits and  
513 computation, pp. 701–708.
- 514 Chandrasekaran, Chandramouli, Illiana E. Bray, and Krishna V. Shenoy (2019). "Frequency Shifts and Depth  
515 Dependence of Premotor Beta Band Activity during Perceptual Decision-Making". *Journal of Neuroscience* 39.8,  
516 pp. 1420–1435.
- 517 Chandrasekaran, Chandramouli, Joana Soldado-Magraner, Diogo Peixoto, William T Newsome, Krishna Shenoy,  
518 and Maneesh Sahani (2018). "Brittleness in model selection analysis of single neuron firing rates". *bioRxiv*.
- 519 Churchland, Mark M., John P. Cunningham, Matthew T. Kaufman, Justin D. Foster, Paul Nuyujukian, Stephen I. Ryu,  
520 and Krishna V. Shenoy (2012). "Neural population dynamics during reaching". *Nature* 487, 51 EP.
- 521 Cury, Kevin M. and Naoshige Uchida (2010). "Robust Odor Coding via Inhalation-Coupled Transient Activity in the  
522 Mammalian Olfactory Bulb". *Neuron* 68.3, pp. 570–585.
- 523 Cuturi, Marco and Mathieu Blondel (2017). "Soft-DTW: a Differentiable Loss Function for Time-Series". *Proceedings  
524 of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70.  
525 Proceedings of Machine Learning Research. International Convention Centre, Sydney, Australia: PMLR, pp. 894–  
526 903.
- 527 Denève, Sophie and Christian K Machens (2016). "Efficient codes and balanced networks". *Nature Neuroscience*  
528 19, p. 375.
- 529 Dhawale, Ashesh K, Rajesh Poddar, Steffen BE Wolff, Valentin A Normand, Evi Kopelowitz, and Bence P Ölveczky  
530 (2017). "Automated long-term recording and analysis of neural activity in behaving animals". *eLife* 6. Ed. by  
531 Andrew J King, e27702.
- 532 Duncker, Lea and Maneesh Sahani (2018). "Temporal alignment and latent Gaussian process factor inference in  
533 population spike trains". *bioRxiv*.
- 534 Feierstein, Claudia E., Michael C. Quirk, Naoshige Uchida, Dara L. Sosulski, and Zachary F. Mainen (2006).  
535 "Representation of Spatial Goals in Rat Orbitofrontal Cortex". *Neuron* 51.4, pp. 495–507.

- 536 Fries, Pascal, Pieter R. Roelfsema, Andreas K. Engel, Peter König, and Wolf Singer (1997). "Synchronization  
537 of oscillatory responses in visual cortex correlates with perception in interocular rivalry". *Proceedings of the  
538 National Academy of Sciences* 94.23, pp. 12699–12704.
- 539 Gao, Peiran and Surya Ganguli (2015). "On simplicity and complexity in the brave new world of large-scale  
540 neuroscience". *Curr Opin Neurobiol* 32. Large-Scale Recording Technology (32), pp. 148–155.
- 541 Gilja, Vikash, Paul Nuyujukian, Cindy A Chestek, John P Cunningham, Byron M Yu, Joline M Fan, Mark M  
542 Churchland, Matthew T Kaufman, Jonathan C Kao, Stephen I Ryu, and Krishna V Shenoy (2012). "A high-  
543 performance neural prosthesis enabled by control algorithm design". *Nature Neuroscience* 15, p. 1752.
- 544 Gollisch, Tim (2006). "Estimating receptive fields in the presence of spike-time jitter". *Network: Computation in  
545 Neural Systems* 17.2, pp. 103–129.
- 546 Goris, Robbe L T, J Anthony Movshon, and Eero P Simoncelli (2014). "Partitioning neuronal variability". *Nat Neurosci*  
547 17.6, pp. 858–865.
- 548 Grosenick, Logan, Brad Klingenberg, Kiefer Katovich, Brian Knutson, and Jonathan E Taylor (2013). "Interpretable  
549 whole-brain prediction analysis with GraphNet". *NeuroImage* 72, pp. 304–321.
- 550 Harvey, Christopher D., Philip Coen, and David W. Tank (2012). "Choice-specific sequences in parietal cortex during  
551 a virtual-navigation decision task". *Nature* 484.7392, pp. 62–68.
- 552 Hill, Daniel N., John C. Curtis, Jeffrey D. Moore, and David Kleinfeld (2011). "Primary Motor Cortex Reports Efferent  
553 Control of Vibrissa Motion on Multiple Timescales". *Neuron* 72.2, pp. 344–356.
- 554 Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment". *Computing In Science & Engineering* 9.3, pp. 90–95.
- 555 Jarvis, M. R. and P. P. Mitra (2001). "Sampling Properties of the Spectrum and Coherency of Sequences of Action  
556 Potentials". *Neural Computation* 13.4, pp. 717–749.
- 557 Jazayeri, Mehrdad and Michael N. Shadlen (2015). "A Neural Mechanism for Sensing and Reproducing a Time  
558 Interval". *Current Biology* 25.20, pp. 2599–2609.
- 559 Jones, Eric, Travis Oliphant, Pearu Peterson, et al. (2001–). *SciPy: Open source scientific tools for Python*.
- 560 Kaufman, Matthew T., Jeffrey S. Seely, David Sussillo, Stephen I. Ryu, Krishna V. Shenoy, and Mark M. Churchland  
561 (2016). "The Largest Response Component in the Motor Cortex Reflects Movement Timing but Not Movement  
562 Type". *eNeuro* 3.4.
- 563 Kawai, Risa, Timothy Markman, Rajesh Poddar, Raymond Ko, Antoniu L. Fantana, Ashesh K. Dhawale, Adam R.  
564 Kampff, and Bence P. Ölveczky (2015). "Motor Cortex Is Required for Learning but Not for Executing a Motor  
565 Skill". *Neuron* 86.3, pp. 800–812.
- 566 Khanna, Preeya and Jose M Carmena (2017). "Beta band oscillations in motor cortex reflect neural population  
567 signals that delay movement onset". *eLife* 6. Ed. by Pascal Fries, e24573.
- 568 — (2015). "Neural oscillations: beta band activity across motor networks". *Current Opinion in Neurobiology* 32.  
569 Large-Scale Recording Technology (32), pp. 60–67.
- 570 Kobak, Dmitry, Wieland Brendel, Christos Constantinidis, Claudia E Feierstein, Adam Kepecs, Zachary F Mainen,  
571 Xue-Lian Qi, Ranulfo Romo, Naoshige Uchida, and Christian K Machens (2016). "Demixed principal component  
572 analysis of neural population data". *eLife* 5. Ed. by Mark CW van Rossum, e10989.
- 573 Krakauer, John W., Asif A. Ghazanfar, Alex Gomez-Marin, Malcolm A. MacIver, and David Poeppel (2017). "Neuro-  
574 science Needs Behavior: Correcting a Reductionist Bias". *Neuron* 93.3, pp. 480–490.
- 575 Lam, Siu Kwan, Antoine Pitrou, and Stanley Seibert (2015). "Numba: A LLVM-based Python JIT Compiler".  
576 *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*. LLVM '15. Austin, Texas:  
577 ACM, 7:1–7:6.
- 578 Latimer, Kenneth (2018). *Simplicity Is Good, But Only When It's Correct*. <http://www.kennethlatimer.com/blog/simplicity-is-good-but-only-when-its-correct>. Blog.
- 579 Lawlor, Patrick N, Matthew G Perich, Lee E Miller, and Konrad P Kording (2018). "Linear-Nonlinear-Time-Warp-  
580 Poisson models of neural activity". *bioRxiv*.

- 582 Leonardo, Anthony and Michale S. Fee (2005). "Ensemble Coding of Vocal Control in Birdsong". *J Neurosci* 25.3,  
583 pp. 652–661.
- 584 London, Michael, Arnd Roth, Lisa Beeren, Michael Häusser, and Peter E Latham (2010). "Sensitivity to perturbations  
585 in vivo implies high noise and suggests rate coding in cortex". *Nature* 466, p. 123.
- 586 Macke, J. H., L. Buesing, and M. Sahani (2015). "Estimating state and parameters in state space models of spike  
587 trains". *Advanced State Space Methods for Neural and Clinical Data*. Ed. by ZheEditor Chen. Cambridge  
588 University Press, pp. 137–159.
- 589 Mackevicius, Emily L, Andrew H Bahle, Alex H Williams, Shijie Gu, Natalia I Denisenko, Mark S Goldman, and  
590 Michale S Fee (2019). "Unsupervised discovery of temporal sequences in high-dimensional datasets, with  
591 applications to neuroscience". *eLife* 8. Ed. by Laura Colgin and Timothy E Behrens, e38471.
- 592 Maheswaranathan, Niru, David B. Kastner, Stephen A. Baccus, and Surya Ganguli (2018). "Inferring hidden structure  
593 in multilayered neural circuits". *PLOS Computational Biology* 14.8, pp. 1–30.
- 594 Marron, J. S., James O. Ramsay, Laura M. Sangalli, and Anuj Srivastava (2015). "Functional Data Analysis of  
595 Amplitude and Phase Variation". *Statist. Sci.* 30.4, pp. 468–484.
- 596 Mueen, Abdullah and Eamonn Keogh (2016). "Extracting Optimal Performance from Dynamic Time Warping".  
597 *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.  
598 KDD '16. San Francisco, California, USA: ACM, pp. 2129–2130.
- 599 Murakami, Masayoshi, M Inês Vicente, Gil M Costa, and Zachary F Mainen (2014). "Neural antecedents of  
600 self-initiated actions in secondary motor cortex". *Nature Neuroscience* 17, p. 1574.
- 601 Murthy, V N and E E Fetz (1992). "Coherent 25- to 35-Hz oscillations in the sensorimotor cortex of awake behaving  
602 monkeys." *Proceedings of the National Academy of Sciences* 89.12, pp. 5670–5674.
- 603 Opsomer, Jean, Yuedong Wang, and Yuhong Yang (2001). "Nonparametric Regression with Correlated Errors".  
604 *Statistical Science* 16.2, pp. 134–153.
- 605 Owen, Art B. and Patrick O. Perry (2009). "Bi-cross-validation of the SVD and the nonnegative matrix factorization".  
606 *Ann. Appl. Stat.* 3.2, pp. 564–594.
- 607 Pandarinath, Chethan, Daniel J. O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D. Stavisky, Jonathan C.  
608 Kao, Eric M. Trautmann, Matthew T. Kaufman, Stephen I. Ryu, Leigh R. Hochberg, Jaimie M. Henderson,  
609 Krishna V. Shenoy, L. F. Abbott, and David Sussillo (2018). "Inferring single-trial neural population dynamics  
610 using sequential auto-encoders". *Nature Methods* 15.10, pp. 805–815.
- 611 Paninski, Liam (2004). "Maximum likelihood estimation of cascade point-process neural encoding models". *Network:  
612 Computation in Neural Systems* 15.4, pp. 243–262.
- 613 Perez, Oswaldo, Robert E. Kass, and Hugo Merchant (2013). "Trial time warping to discriminate stimulus-related  
614 from movement-related neural activity". *Journal of Neuroscience Methods* 212.2, pp. 203–210.
- 615 Perry, Patrick O. (2009). "Cross-Validation for Unsupervised Learning". PhD thesis. Stanford University.
- 616 Petitjean, François, Alain Ketterlin, and Pierre Gançarski (2011). "A global averaging method for dynamic time  
617 warping, with applications to clustering". *Pattern Recognition* 44.3, pp. 678–693.
- 618 Petreska, Biljana, Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen I. Ryu, Krishna V Shenoy, and  
619 Maneesh Sahani (2011). "Dynamical segmentation of single trials from population neural data". *Advances in  
620 Neural Information Processing Systems* 24. Ed. by J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and  
621 K. Q. Weinberger. Curran Associates, Inc., pp. 756–764.
- 622 Pnevmatikakis, Eftychios A. and Andrea Giovannucci (2017). "NoRMCorre: An online algorithm for piecewise rigid  
623 motion correction of calcium imaging data". *Journal of Neuroscience Methods* 291, pp. 83–94.
- 624 Poddar, Rajesh, Risa Kawai, and Bence P. Ölveczky (2013). "A Fully Automated High-Throughput Training System  
625 for Rodents". *PLOS ONE* 8.12.

- 626 Poole, Ben, Alex H. Williams, Niru Maheswaranathan, Byron Yu, Gopal Santhanam, Stephen Ryu, Stephen A.  
627 Baccuss, Krishna V. Shenoy, and Surya Ganguli (2017). "Time-warped PCA: simultaneous alignment and  
628 dimensionality reduction of neural data". *Cosyne Abstracts*. Salt Lake City, UT, USA.
- 629 Ray, Supratim (2015). "Challenges in the quantification and interpretation of spike-LFP relationships". *Current*  
630 *Opinion in Neurobiology* 31. SI: Brain rhythms and dynamic coordination, pp. 111–118.
- 631 Reimer, Jacob and Nicholas G. Hatsopoulos (2010). "Periodicity and Evoked Responses in Motor Cortex". *J Neurosci*  
632 30.34, pp. 11506–11515.
- 633 Sanes, J N and J P Donoghue (1993). "Oscillations in local field potentials of the primate motor cortex during  
634 voluntary movement." *Proc Natl Acad Sci U S A* 90.10, pp. 4470–4474.
- 635 Shokoohi-Yekta, Mohammad, Jun Wang, and Eamonn Keogh (2015). "On the Non-Trivial Generalization of Dynamic  
636 Time Warping to the Multi-Dimensional Case". *Proceedings of the 2015 SIAM International Conference on Data*  
637 *Mining*, pp. 289–297.
- 638 Shushruth, S., Mark Mazurek, and Michael N. Shadlen (2018). "Comparison of Decision-Related Signals in Sensory  
639 and Motor Preparatory Responses of Neurons in Area LIP". *Journal of Neuroscience* 38.28, pp. 6350–6365.
- 640 Shusterman, Roman, Yevgeniy B. Sirotin, Matthew C. Smear, Yashar Ahmadian, and Dmitry Rinberg (2018). "Sniff  
641 Invariant Odor Coding". *eNeuro* 5.6.
- 642 Shusterman, Roman, Matthew C Smear, Alexei A Koulakov, and Dmitry Rinberg (2011). "Precise olfactory responses  
643 tile the sniff cycle". *Nat Neurosci* 14.8, pp. 1039–1044.
- 644 Smith, Carl and Liam Paninski (2013). "Computing loss of efficiency in optimal Bayesian decoders given noisy or  
645 incomplete spike trains". *Network: Computation in Neural Systems* 24.2, pp. 75–98.
- 646 Softky, WR and C Koch (1993). "The highly irregular firing of cortical cells is inconsistent with temporal integration of  
647 random EPSPs". *Journal of Neuroscience* 13.1, pp. 334–350.
- 648 Starkweather, Clara Kwon, Benedicte M Babayan, Naoshige Uchida, and Samuel J Gershman (2017). "Dopamine  
649 reward prediction errors reflect hidden-state inference across time". *Nature Neuroscience* 20, p. 581.
- 650 Stevenson, Ian H and Konrad P Kording (2011). "How advances in neural recording affect data analysis". *Nature*  
651 *Neuroscience* 14, p. 139.
- 652 Sun, Felice T., Lee M. Miller, and Mark D'Esposito (2005). "Measuring temporal dynamics of functional networks  
653 using phase spectrum of fMRI data". *NeuroImage* 28.1, pp. 227–237.
- 654 Tavenard, Romain (2017). *tslearn: A machine learning toolkit dedicated to time-series data*. <https://github.com/rtavenar/tslearn>.
- 655 Tort, Adriano B L, Simon Ponsel, Jakob Jessberger, Yevgenij Yanovsky, Jurij Brankač, and Andreas Draguhn  
656 (2018). "Parallel detection of theta and respiration-coupled oscillations throughout the mouse brain". *Scientific*  
657 *Reports* 8.1, p. 6432.
- 658 Ventura, Valérie (2004). "Testing for and estimating latency effects for Poisson and non-Poisson spike trains". *Neural*  
659 *Computation* 16.11, pp. 2323–2349.
- 660 Wang, Jing, Devika Narain, Eghbal A. Hosseini, and Mehrdad Jazayeri (2018). "Flexible timing by temporal scaling  
661 of cortical responses". *Nature Neuroscience* 21.1, pp. 102–110.
- 662 Williams, Alex H., Tony Hyun Kim, Forea Wang, Saurabh Vyas, Stephen I. Ryu, Krishna V. Shenoy, Mark Schnitzer,  
663 Tamara G. Kolda, and Surya Ganguli (2018). "Unsupervised Discovery of Demixed, Low-Dimensional Neural  
664 Dynamics across Multiple Timescales through Tensor Component Analysis". *Neuron* 98.6, 1099–1115.e8.
- 665 Wilson, Christopher D., Gabriela O. Serrano, Alexei A. Koulakov, and Dmitry Rinberg (2017). "A primacy code for  
666 odor identity". *Nature Communications* 8.1, p. 1477.
- 667 Wright, Stephen J. (2015). "Coordinate descent algorithms". *Mathematical Programming* 151.1, pp. 3–34.
- 668 Yger, Pierre, Giulia LB Spampinato, Elric Esposito, Baptiste Lefebvre, StÃ©phane Deny, Christophe Gardella,  
669 Marcel Stimberg, Florian Jetter, Guenther Zeck, Serge Picaud, Jens Duebel, and Olivier Marre (2018). "A spike

- 671 sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo". *eLife*  
672 7. Ed. by David Kleinfeld, e34518.  
673 Zhang, Zheng, Romain Tavenard, Adeline Bailly, Xiaotong Tang, Ping Tang, and Thomas Corpetti (2017). "Dynamic  
674 Time Warping under limited warping path length". *Information Sciences* 393, pp. 91–107.

## 675 Methods

### 676 Code Implementation & Availability

677 Our code for fitting linear and piecewise linear time warping models is distributed as a GitHub repository (under  
678 an MIT license): <https://github.com/ahwillia/affinewarp>. Our Python implementation relies on the standard  
679 SciPy scientific computing libraries (Jones et al. 2001–; Hunter 2007). Additionally, we achieved substantial  
680 performance enhancements by leveraging `numba`, a Python library that enables just-in-time (JIT) compilation (Lam  
681 et al. 2015). Step-by-step tutorials for executing our code are available on GitHub.

### 682 Detailed Description of Time Warping

#### 683 Notation

684 We follow the same notation introduced in the main text. Matrices are denoted in bold, uppercase fonts, e.g.  $\mathbf{M}$ ,  
685 while vectors are denoted in bold, lowercase fonts, e.g.  $\mathbf{v}$ . Unless otherwise specified, non-boldface letters specify  
686 scalar quantities, e.g.  $S$  or  $s$ . We use  $\mathbf{M}^T$  and  $\mathbf{M}^{-1}$  to denote the transpose and inverse of a matrix, respectively.

687 We consider a dataset consisting of  $N$  features over  $K$  trials with  $T$  timesteps per trial. For simplicity, we refer to  $N$   
688 as the number of neurons in the dataset; however,  $N$  could also refer to the number of fMRI voxels, multiunits, or  
689 regions of interest in imaging data. The full dataset is a third-order tensor (a three-dimensional data array) with  
690 dimensions  $K \times T \times N$ . The  $k^{\text{th}}$  slice of the data tensor is a  $T \times N$  matrix  $\mathbf{X}_k$ , which represents the activity of  
691 the neural population on trial  $k$ . We denote a single element of the tensor as  $X_{k,t,n}$ , which specifies the activity of  
692 neuron  $n$  at timebin  $t$  on trial  $k$ .

693 The time warping model produces an estimate of population activity on each trial. Mirroring standard notation in  
694 linear regression, we denote the model estimate on trial  $k$  as  $\hat{\mathbf{X}}_k$  (a  $T \times N$  matrix).

#### 695 Model Estimate and Template Interpolation Scheme

696 The main idea behind time warping is to approximate each trial,  $\mathbf{X}_k$ , as a warped version of a  $N \times T$  template,  $\tilde{\mathbf{X}}$ ,  
697 that is shared across all trials. For neuron  $n$ , at time bin  $t$ , on trial  $k$ , the spirit behind the model is:

$$\hat{X}_{k,t,n} = \tilde{X}_{\omega_k(t),n} \quad (3)$$

698 However, this expression is only valid when the warping function,  $\omega_k(t)$ , produces integer values. To allow the  
699 warping functions to produce non-integer values, we adopt a standard linear interpolation scheme. Let  $\omega_k : t \mapsto \tau$   
700 describe the time warping function for trial  $k$ , such that  $t$  is the integer-valued time index for the data (clock time),  
701 and  $\tau$  is any real number representing time for the response template. Then, the model estimate for neuron  $n$ , at  
702 time bin  $t$ , on trial  $k$  is given by:

$$\hat{X}_{k,t,n} = (\lceil \tau \rceil - \tau) \tilde{X}_{\lfloor \tau \rfloor, n} + (\tau - \lceil \tau \rceil) \tilde{X}_{\lceil \tau \rceil, n} \quad (4)$$

703 where  $\tau = \omega_k(t)$ ,  $\lfloor \cdot \rfloor$  represents the “flooring” operation, and  $\lceil \cdot \rceil$  represents the “ceiling” operation. Note that  $\tau$   
704 implicitly depends on the trial index  $k$ , but we do not explicitly denote this dependence for notational simplicity.

Because the model estimate (**Eq 4**) is a linear combination of  $\tilde{X}_{\lfloor \tau \rfloor n}$  and  $\tilde{X}_{\lceil \tau \rceil n}$ , the warping transformation can be represented as a matrix  $\mathbf{W}$  with elements:

$$\begin{aligned} W_{t,\lceil \tau \rceil} &= \tau - \lfloor \tau \rfloor \\ W_{t,\lfloor \tau \rfloor} &= \lceil \tau \rceil - \tau \end{aligned} \quad (5)$$

For each trial, the *warping matrix*  $\mathbf{W}_k$  can be uniquely determined from the warping function  $\omega_k$ . Thus, the model estimate on each trial is given by:

$$\hat{\mathbf{X}}_k = \mathbf{W}_k \tilde{\mathbf{X}} \quad (6)$$

## Optimization Strategy

The model template and warping functions are optimized to minimize an objective function, which we denote as  $F(\tilde{\mathbf{X}}, \omega_1, \omega_2, \dots, \omega_K)$ . We assume that this objective function decomposes across trials as follows:

$$F(\tilde{\mathbf{X}}, \omega_1, \omega_2, \dots, \omega_K) = \sum_{k=1}^K f_k(\tilde{\mathbf{X}}, \omega_k) + \rho_1(\tilde{\mathbf{X}}) \quad (7)$$

Here  $f_k$  is a function defining the model loss on trial  $k$ , and  $\rho_1$  is a regularization term, penalizing the roughness and size of the template (described in the next section). Our online code package supports least-squares and Poisson loss functions; we adopted the least-squares criterion for the purposes of this paper due to its computational efficiency and its ability to be adapted to non-spike time data (e.g. fMRI or calcium imaging). Under this choice, the per-trial loss function is:

$$f_k(\tilde{\mathbf{X}}, \omega_k) = \|\mathbf{W}_k \tilde{\mathbf{X}} - \mathbf{X}_k\|_F^2 + \rho_2(\omega_k) \quad (8)$$

Here,  $\rho_2$  is a regularization term that penalizes the magnitude of warping (described in the next section), and  $\|\cdot\|_F^2$  denotes the squared Frobenius norm, which is simply the sum of squared residuals,  $\|\mathbf{M}\|_F^2 = \sum_{ij} M_{ij}^2$ .

To minimize  $F$ , we adopt an alternating optimization (block coordinate descent) approach (Wright 2015). First, each warping function is initialized to be the identity,  $\omega_k(t) = t$ , and the template and warping functions are cyclically updated according to the following sequence of optimization subproblems:

$$\begin{aligned} \tilde{\mathbf{X}} &\leftarrow \underset{\tilde{\mathbf{X}}}{\operatorname{argmin}} F(\tilde{\mathbf{X}}, \underline{\omega}_1, \dots, \underline{\omega}_K) \\ \omega_1 &\leftarrow \underset{\underline{\omega}_1}{\operatorname{argmin}} F(\tilde{\mathbf{X}}, \underline{\omega}_1, \dots, \underline{\omega}_K) \\ &\vdots \\ \omega_K &\leftarrow \underset{\underline{\omega}_K}{\operatorname{argmin}} F(\tilde{\mathbf{X}}, \underline{\omega}_1, \dots, \underline{\omega}_K) \end{aligned} \quad (9)$$

Here, an underlined variable denotes a dummy variable that is optimized over in each subproblem. This sequence of parameter updates is cyclically repeated until the objective value ceases to improve; by construction, the objective monotonically decreases at each step of the algorithm so convergence is guaranteed under mild assumptions (Wright 2015).

This partitioning the parameter updates enables each subproblem to be solved very efficiently. When the template is considered a fixed variable, the objective function decouples across trials (**Eq 7**), which simplifies the warping function updates considerably:

$$\omega_k \leftarrow \underset{\underline{\omega}_k}{\operatorname{argmin}} f_k(\tilde{\mathbf{X}}, \omega_k) \quad (10)$$

729 These parameter updates are entirely independent, with each update only depending on the raw data for trial  $k$ ,  $\mathbf{X}_k$ ,  
 730 and the current warping template  $\tilde{\mathbf{X}}$ . Our code package executes them efficiently in parallel across CPU threads.  
 731 Furthermore, each warping function is controlled by a small number of parameters in our framework—at best a  
 732 single parameter (shift-only warping) and at worst only a few parameters (piecewise linear warping). Thus, we  
 733 perform these updates by a brute force random search (see *Warping Function Regularization and Update Rule*).  
 734 The response template is also very simple to update, especially under a least-squares loss criterion. Assume for  
 735 the moment that the model is not regularized; i.e.,  $\rho_1(\tilde{\mathbf{X}}) = 0$  and  $\rho_2(\mathbf{W}_k) = 0$ . Then, because each  $\mathbf{W}_k$  is held  
 736 constant, updating the template amounts to a least-squares problem that can be solved in closed form:

$$\begin{aligned} \operatorname{argmin}_{\tilde{\mathbf{X}}} \sum_{k=1}^K \|\mathbf{W}_k \tilde{\mathbf{X}} - \mathbf{X}_k\|_F^2 &= \operatorname{argmin}_{\tilde{\mathbf{X}}} \left\| \left( \sum_k \mathbf{W}_k^\top \mathbf{W}_k \right) \tilde{\mathbf{X}} - \left( \sum_k \mathbf{W}_k^\top \mathbf{X}_k \right) \right\|_F^2 \\ &= \left( \sum_k \mathbf{W}_k^\top \mathbf{W}_k \right)^{-1} \sum_k \mathbf{W}_k^\top \mathbf{X}_k \end{aligned} \quad (11)$$

737 Furthermore the matrix  $\sum_{k=1}^K \mathbf{W}_k \mathbf{W}_k^\top$  is a *symmetric, tridiagonal matrix*. Intuitively, the tridiagonal structure arises  
 738 from the constraint that each warping function is monotonically increasing, and the local structure of the linear  
 739 interpolation scheme. Consider any warping matrix  $\mathbf{W}$ , with an associated warping function  $\omega$ . **Equation 4** implies  
 740 that  $W_{i,t} W_{j,t} = 0$  if  $|\omega(t) - i| > 1$  or if  $|\omega(t) - j| > 1$  and thus  $[\mathbf{W}^\top \mathbf{W}]_{i,j} = 0$  if  $|i - j| > 1$ .

741 The tridiagonal structure of  $\sum_k \mathbf{W}_k^\top \mathbf{W}_k$  enables the template parameters to be updated extraordinarily fast for  
 742 practical applications. We use a specialized solver for systems of linear equations with banded, symmetric structure  
 743 (`scipy.linalg.solveh_banded`). This allows the matrix inversion in **Equation 11** to be (implicitly) carried out in  
 744  $O(TN)$  operations instead of  $O(T^3 + T^2N)$  operations if  $\mathbf{W}$  was treated as a dense matrix.

## 745 Template Regularization and Update Rule

746 We found that introducing regularization (penalties on the magnitude or complexity of model parameters) can  
 747 improve the interpretability of the model and its ability to predict held out data. First, we found in some datasets  
 748 that the warping template could exhibit rapid, high-frequency changes in firing rate (see, e.g., the template in **Fig 1C**, which was fit without regularization). These irregularities likely correspond to the model overfitting to noisy  
 749 neuronal data, and can be discouraged by penalizing the magnitude of the second finite differences along the  
 750 temporal dimension of the template (Grosenick et al. 2013; Maheswaranathan et al. 2018). We refer to this term as  
 751 a *roughness penalty* or *smoothness regularization*. Second, it is possible that the matrix  $\sum_k \mathbf{W}_k^\top \mathbf{W}_k$  appearing in  
 752 eq. (11) would become non-invertible or ill-conditioned during optimization. To prevent this, and to discourage the  
 753 template firing rates from becoming too large, we added a penalty on the squared Frobenius norm of the template.  
 754 Formally, the regularization on the template is given by:

$$\rho_1(\tilde{\mathbf{X}}) = \lambda \left\| \mathbf{D} \tilde{\mathbf{X}} \right\|_F^2 + \gamma \left\| \tilde{\mathbf{X}} \right\|_F^2 \quad (12)$$

756 where  $\lambda > 0$  controls the strength of the roughness penalty and  $\gamma > 0$  controls the strength of the Frobenius norm  
 757 penalty. The matrix  $\mathbf{D}$  is a  $(T - 2) \times T$  matrix that computes second-order finite differences:

$$\mathbf{D} = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & & \vdots \\ \vdots & & & \ddots & & 0 \\ 0 & \dots & 0 & 1 & -2 & 1 \end{bmatrix} \quad (13)$$

758 Incorporating this regularization term into the update of the warping template (eq. (11)), we get:

$$\begin{aligned} \operatorname{argmin}_{\tilde{\mathbf{X}}} & \left\| \left( \sum_k \mathbf{W}_k^T \mathbf{W}_k \right) \tilde{\mathbf{X}} - \left( \sum_k \mathbf{W}_k^T \mathbf{X}_k \right) \right\|_F^2 + \lambda \left\| \mathbf{D} \tilde{\mathbf{X}} \right\|_F^2 + \gamma \left\| \tilde{\mathbf{X}} \right\|_F^2 \\ & = \operatorname{argmin}_{\tilde{\mathbf{X}}} \left\| \left( \sum_k \mathbf{W}_k^T \mathbf{W}_k + \lambda \mathbf{D}^T \mathbf{D} + \gamma \mathbf{I} \right) \tilde{\mathbf{X}} - \left( \sum_k \mathbf{W}_k^T \mathbf{X}_k \right) \right\|_F^2 \end{aligned} \quad (14)$$

759 Which yields the template update rule:

$$\tilde{\mathbf{X}} \leftarrow \left( \sum_k \mathbf{W}_k^T \mathbf{W}_k + \lambda \mathbf{D}^T \mathbf{D} + \gamma \mathbf{I} \right)^{-1} \sum_k \mathbf{W}_k^T \mathbf{X}_k \quad (15)$$

760 Thus, the solution is the same as before except a term  $\lambda \mathbf{D}^T \mathbf{D} + \gamma \mathbf{I}$  is added to the inverted matrix (left-hand side  
 761 of linear system). These modifications hardly affect the computational complexity of the parameter update since  
 762  $\lambda \mathbf{D}^T \mathbf{D} + \gamma \mathbf{I}$  is also a symmetric, banded matrix. Furthermore, as long as  $\gamma > 0$  the overall matrix is positive definite  
 763 and therefore guaranteed to be invertible.

764 In practice, we have found that it is simple to hand-tune the regularization strengths for exploratory analysis (though  
 765 cross-validation procedures, described below, should always be used to monitor for overfitting). We typically set  
 766 the L2 regularization ( $\gamma$ ) to be zero or very small (e.g., 1e-4) and do not tune it further. A reasonable value for the  
 767 roughness penalty scale can be found by visually inspecting the template for various neurons (columns of  $\tilde{\mathbf{X}}$ ) and  
 768 increasing  $\lambda$  if these time series appear noisy.

## 769 Warping Function Regularization and Update Rule

770 We found that the optimization landscape of linear and piecewise linear warping functions is complex and full of  
 771 local minima. Thus, gradient-based optimization methods can be ineffective. Thankfully, the warping functions are  
 772 (a) low-dimensional and (b) entirely decoupled across trials. Thus, when updating the warping functions, we perform  
 773 a brute force parameter search for each trial in parallel. For shift-only warping models, we perform a dense grid  
 774 search over the parameter (the magnitude of the shift).

775 For piecewise linear warping models we perform an annealed random search as follows. Consider a warping  
 776 function  $\omega(t)$  for any arbitrary trial (we drop the trial index  $k$  for brevity). We parameterize the warping function as:

$$\omega(t) = 1 + (T - 1) \cdot \left[ \tilde{\omega} \left( \frac{t - 1}{T - 1} \right) \right]_0^1 \quad (16)$$

777 where  $\tilde{\omega}$  is a piecewise linear function mapping the unit interval  $[0, 1]$  to any real number, and  $[z]_0^1 = \max(\min(z, 1), 0)$   
 778 denotes clipping any real number  $z$  to have a value between zero and one.

779 The piecewise linear function  $\tilde{\omega}$  is defined by a series of  $M$  x-y coordinates,  $\{(\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_M, \beta_M)\}$ , where  
 780  $0 = \alpha_1 < \alpha_2 < \dots < \alpha_M = 1$  and  $\beta_1 \leq \beta_2 \leq \dots \leq \beta_M$ . We refer to these coordinates as the *knots* of the warping

781 function. The function is defined using linear interpolation:

$$\tilde{\omega}(t) = \begin{cases} \beta_1 & \text{if } t = \alpha_1 = 0 \\ \beta_1 \left(1 - \frac{t-\alpha_1}{\alpha_2-\alpha_1}\right) + \beta_2 \left(\frac{t-\alpha_1}{\alpha_2-\alpha_1}\right) & \text{if } \alpha_1 < t < \alpha_2 \\ \vdots & \vdots \\ \beta_{M-1} \left(1 - \frac{t-\alpha_{M-1}}{\alpha_M-\alpha_{M-1}}\right) + \beta_M \left(\frac{t-\alpha_{M-1}}{\alpha_M-\alpha_{M-1}}\right) & \text{if } \alpha_{M-1} < t < \alpha_M \\ \beta_M & \text{if } t = \alpha_M = 1 \end{cases} \quad (17)$$

To optimize the warping functions we perform a random search over these coordinates/knots. Let  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_M]$  and  $\beta = [\beta_1, \beta_2, \dots, \beta_M]$  denote the current coordinates. We form a new proposed warping function by:

$$\text{??}\alpha' \leftarrow \text{sort}(\alpha + Q\eta) \quad (18)$$

$$\alpha' \leftarrow (\alpha' - \alpha'_1) / (\alpha'_M - \alpha'_1) \quad (19)$$

$$\beta' \leftarrow \text{sort}(\beta + Q\eta) \quad (20)$$

where  $Q > 0$  is a scalar parameter tuning the amount of exploration, and  $\eta$  is a vector of random normal variables with mean zero and unit variance. The procedure “sort(v)” re-orders the elements of a vector so that they are in ascending order. If the proposed warping function improves the objective function, we accept the new parameters:

$$\alpha \leftarrow \alpha'$$

$$\beta \leftarrow \beta'$$

782 For each round of optimization we exponentially relax  $Q$  from 1.0 to 0.01 over a fixed number of iterations.

783 We also found that penalizing the warping functions based on their distance from the identity line was helpful in  
 784 some cases. Intuitively, this encourages the warping functions to be minimal—as the penalty strength increases the  
 785 warping functions will approach  $\omega(t) = t$ , resulting in no warping at all in this extreme limit. Similar penalties or hard  
 786 constraints on time warping have been examined in prior literature (see e.g., Zhang et al. 2017). We chose the  
 787 penalty to be the area between the unclipped warping function and the identity line:

$$\rho_2(\mathbf{W}_k) = \mu \int_0^1 |\tilde{\omega}_k(t) - t| dt \quad (21)$$

788 which, for piecewise linear functions with relatively small  $M$ , can be efficiently computed as the sum of triangular and  
 789 trapezoidal regions. Here,  $\mu \geq 0$  is a scalar hyperparameter controlling the strength of the penalty. In practice we  
 790 start with  $\mu = 0$  and increase it if, upon visual inspection, the warping functions are highly deviant from the identity  
 791 line. Increasing  $\mu$  in these cases can result in more sensible and interpretable templates. Again, cross-validation  
 792 procedures can be used to assess whether  $\mu$  is too low (resulting in overfitting) or too high (resulting in underfitting).

### 793 Cross-validation

794 As with any statistical method, one must be very careful that time warping does not reveal spurious structure and  
 795 features of the data. In **Fig 1**, we saw that even a simple linear warping model can result in noticeable overfitting on a

simple synthetic time series. An important technical contribution of our work is a rigorous cross-validation framework for time warping models. This framework, described in detail below, enables us to fine-tune all regularization terms—i.e. the hyperparameters  $\{\gamma, \lambda, \mu\}$ —across all warping models. That is, we can rigorously compare the performance of shift-only, linear, and piecewise-linear time warping models on an even footing, and thus critically examine the degree of nonlinearity in time warping. While cross-validation is a common procedure in statistical modeling and in modern neuroscience, there are subtle pitfalls that must be avoided in unsupervised learning models (Bro et al. 2008; Perry 2009), and in models with smoothness regularization terms (Opsomer et al. 2001). These concerns are not merely theoretical; they have directly impacted recent results in neuroscience (Latimer 2018).

To properly compare the performance of different warping models, it is important to perform *nested* cross-validation, so that regularization terms are separately tuned for each model. For example, a piecewise linear warping model will often require stronger smoothness and warp regularization terms, compared to a simpler, shift-only warping model. Thus, on each cross-validation run we split the data in three partitions: a training set, a validation set, and a test set. For each model class (shift-only, linear warping, piecewise-linear warping, etc.) we fit 100 models with randomized regularization strengths to the training set; we then evaluated all 100 models on the held-out validation set; finally, the best-performing model was evaluated on the test set. The test set performance is then compared across model classes. We used  $\sim 73\%$  of the data for training,  $\sim 13\%$  for validation, and the final  $\sim 13\%$  for testing. We performed this overall cross-validation procedure 100 times, drawing different randomized data partitions each time—this is known as *randomized cross-validation* and is useful for heterogeneous datasets, in which features (i.e. neurons) exhibit varied levels of noise.

Recall that our dataset consists of  $N$  neurons,  $T$  timebins, and  $K$  trials. The question then arises, should we hold out neurons, timepoints, or trials during cross-validation? Since the warping functions are assumed to be shared across all neurons, these model parameters can be fit on a subset of neurons (training set), and then evaluated on held out neurons (validation/test sets). However, if we hold out individual neurons entirely, then it is impossible to fit the response template matrix  $\tilde{X}$  for those cells. Conversely, the response template can be fit to a subset of trials (training set) and evaluated on the remaining trials (validation/test sets). However, if we hold out individual trials entirely, then it is impossible to fit the warping functions associated with those held out trials.

To circumvent this problem we adopt a bi-cross-validation hold out pattern (Owen and Perry 2009). This entails separately and independently partitioning neurons and trials. Thus, we randomly choose training neurons ( $\sim 73\%$  of all cells), validation neurons ( $\sim 13\%$  of all cells), and testing neurons (the remaining  $\sim 13\%$ ). Additionally, we randomly choose training trials, validation trials, and testing trials, according to these same ratios. The model warping functions are fit to all trials, but only on the training neurons; the response template is fit for all neurons, but only on the training trials. When reporting the training performance, we compute the reconstruction loss on the *intersection* of the training neurons and training trials. Likewise, when evaluating models on the validation (or test) set, we compute the the reconstruction loss on the intersection of the validation (or test) neurons and validation (or test) trials.

Temporal dependencies in model errors can complicate proper cross-validation (Opsomer et al. 2001; Latimer 2018). To avoid these complications, we leave out entire trials for individual neurons, rather than leaving out a subset of time bins.

### Null models and other sanity checks

The cross-validation procedure described above is fully rigorous, but computationally expensive to perform. Even if each optimization run only takes a few seconds to complete, comparing  $M$  warping models over  $P$  random samples

838 of the regularization parameters, and repeating the whole process over  $Q$  randomized folds leads to long run times;  
839 for example, ~26 hours for  $M = 5$ ,  $P = 100$ ,  $Q = 50$  and each model taking ~20 seconds to optimize. This rather  
840 unfavorable scaling underscores why our attention to performance enhancements—e.g. by exploiting banded  
841 matrix structure when updating the model template—is critical for practical applications. On the other hand, a full  
842 cross-validation run is often unnecessary for exploratory data analysis and visualization. Here we briefly outline two  
843 simple procedures for validating time warping visualizations in a more interactive manner. Our online code package  
844 also supports both of these options.

845 First, one can create a very simple null dataset of neural activity that, by construction, contains no warping. By  
846 comparing the results of time warping on this null dataset to those achieved on the real data, we gain an informative  
847 reference point. For spiking data, we simulate null data by computing the trial-average firing rate of each neuron  
848 and then drawing Poisson i.i.d. random spike trains on every trial. That is, on each trial, the spike train for a neuron  
849 is drawn from an inhomogeneous Poisson process, with a rate function given by the trial-average firing rate. Similar  
850 baselines could be developed for calcium imaging and fMRI studies after specifying an appropriate noise model.

851 Second, a key visualization tool enabled by time warping is the alignment of neural activity across trials. This  
852 alignment is achieved by applying the inverse warping functions to re-scale the time axis on the raw data; it does *not*  
853 directly rely on the response template,  $\tilde{X}$ . Thus, one can visualize the aligned activity of an individual neuron in a  
854 held out manner—the model is fit to all trials and all other neurons, and the warping functions are applied to the  
855 held out cell. This can then be repeated for each neuron in the full population. All spike raster plots in the main  
856 paper were produced using this procedure.

857 While these two approaches do not supplant the need for careful cross-validation, they can provide a quick validation  
858 for visualizations and presented results.

## 859 Synthetic data examples

860 In **Figure 1** data from a single neuron was simulated as a difference of two exponential curves. The activity at  
861  $T = 100$  equally spaced time points between  $[-8, +8]$  was given by:

$$x_k(t) = \begin{cases} 0 & s_k - t < 0 \\ 3.3 \cdot \exp((s_k - t)/2) - \exp(s_k - t) + \eta & s_k - t \geq 0 \end{cases}$$

862 Where  $s_k$  was a random shift parameter drawn uniformly on the interval  $[-5.5, 3.5]$ , and  $\eta$  was randomly drawn  
863 zero-mean gaussian noise with a standard deviation of 0.15. Unregularized shift-only, linear, and piecewise linear  
864 (with 1 knot) models were fit to  $K = 100$  simulated trials. DTW-Barycenter Averaging (DBA; Petitjean et al. 2011)  
865 was fit to the same data using (Tavenard 2017).

866 In **Figure 2** we simulated random warping warping functions following the procedure listed in **Equation ??**, with  
867  $Q = .12$ . The firing rate template of each neuron was given by a smoothed, sparse sequence of heavy-tailed random  
868 variables:

$$x(t) = 0.01 + \text{conv}(b_t \cdot e_t)$$

869 where  $e_t$  were randomly drawn from an exponential distribution (with scale parameter equal to one) and  $b_t$  were  
870 binary random variables drawn from a Bernoulli distribution (with probability of 0.92 that  $b_t = 0$ ). The  $\text{conv}(\cdot)$   
871 procedure denotes convolution with a Gaussian smoothing kernel with a standard deviation of 2. Truncated Poisson  
872 random variables were then drawn in each timebin; any bins with more than two spikes were truncated to one spike.

## 873 Experimental Methods

### 874 Mouse Olfactory Task

875 All procedures were approved by the Institutional Animal Care and Use Committee of New York University Langone  
876 Medical Center. We analyzed data that was collected as part of a previously published study (Wilson et al. 2017).

877 Data presented from the mouse olfactory bulb were collected from a single recording session using an awake male  
878 C57B/6 mouse. Subject was implanted with a RIVETS headbar for head-fixation 7 days prior to the experiment (see  
879 description in Arneodo et al. 2018). Subjects were water deprived prior to the experiment and were administered  
880 water during random odor presentations to acclimate animals to the experimental apparatus.

881 On the day of experiment, the subject was anesthetized using isoflurane and a ~0.3 mm craniotomy was preformed  
882 to gain access to the dorsal olfactory bulb. NeuroNexus A2x32 probes were inserted approximately 500  $\mu\text{m}$  into  
883 the dorsal bulb to record from the mitral-tufted cell layer. After probe insertion, the subject was allowed to recover  
884 from anesthesia for 30 minutes prior to recording. Electrophysiological and respiration signals were recorded using  
885 the HHMI Janelia Whisper recording system at 25000 Hz. Respiration (sniff) was monitored non-invasively using a  
886 pressure sensor sampling from the airflow in front of the nose. Action potentials from the recording were identified  
887 and classified into units offline using Spyking Circus template-matching software (Yger et al. 2018).

888 Two odors at 3 concentrations were presented in randomly interleaved trials. Subjects were passively sampling  
889 odor during trials. Concentrations covered a range of 2 orders of magnitude of molarity in carrier air. Odorants  
890 were diluted in mineral oil, stored in amber volatile organic analysis vials, and delivered via a 8-odor olfactometer.  
891 Odorant concentrations were controlled using a combination of gas- and liquid-phase dilution. We restricted our  
892 analysis to subsets of trials with odorant concentrations of  $10^{-2}$  M—the highest concentration analyzed in (Wilson  
893 et al. 2017).

### 894 Primate Motor Task

895 All procedures and experiments were reviewed and approved by the Stanford University Institutional Animal Care  
896 and Use Committee. Two male rhesus macaque monkeys (*Macacca mulatta*), denoted monkey J and monkey U,  
897 were used in this study. The monkeys were 13 (J) and 7 (U) years old and weighed 16 kg (J) and 13 kg (U) at the  
898 time of these experiments.

899 Monkeys performed a standard center-out delayed reach task described previously in (Gilja et al. 2012; Ames et al.  
900 2014). Targets were presented at (40, 80, 120) cm and 90 cm from the central starting location for monkeys J and U  
901 respectively. For monkey J, delay periods were evenly distributed between 300 and 700 ms (monkey J) with ~4.5%  
902 non-delay trials randomly interleaved. For monkey U, delays were randomly distributed between 350 and 600 ms on  
903 87% trials, between 5 and 350 ms on 10% of trials, with 3% non-delay trials. Monkeys received a liquid reward upon  
904 touching and holding the cursor on the target. Movement during the delay period caused a trial failure and provided  
905 a brief automated time out (~1 s). Non-delay trials were not analyzed.

906 For monkey J, the virtual cursor and targets were presented in a three-dimensional environment (MusculoSkeletal  
907 Modeling Software, Medical Device Development Facility, University of Southern California). Hand-position data were  
908 measured at 60 Hz with an infrared reflective bead-tracking system (Polaris, Northern Digital). Behavioral control  
909 and neural decode were run on separate PCs using the Simulink/xPC platform (Mathworks) with communication  
910 latencies of less than 3ms. This system enabled millisecond timing precision for all computations. Visual presentation  
911 was provided via two LCD monitors in a Wheatstone stereotax configuration, with refresh rates at 120 Hz, yielding

912 frame updates of  $7 \pm 4$  ms. Two mirrors visually fused the displays into a single three-dimensional percept for the  
913 user, as described previously in ([Gilja et al. 2012](#)).

914 For monkey U, the virtual cursor and targets were presented on a standard 2D display. The monkey controlled the  
915 position of an onscreen cursor using a haptic manipulandum which applied no additional forces applied to the arm  
916 and was only used for positional cursor control. The haptic device was constrained to move within a 2D vertical  
917 workspace and cursor position tracks hand position 1:1 without perceptible lag.

918 Neural recordings were obtained via implanted 96-electrode Utah Microelectrode arrays (Blackrock Microsystems)  
919 using standard neurosurgical techniques. Two arrays were implanted in the left hemisphere of Monkey J, one  
920 in dorsal premotor cortex (PMd) and one in primary motor cortex (M1). Three arrays were implanted in the left  
921 hemisphere of Monkey U, one in PMd, one in medial M1, and one in lateral M1. For both monkeys, implantation  
922 location was estimated visually from local anatomical landmarks.

923 Neural data were band-pass filtered between 250-7500 Hz, and processed to obtain multiunit ‘threshold crossings’  
924 spikes, defined as any time the signal crosses -3.5 times RMS voltage. We did not perform spike sorting, and instead  
925 grouped together the multiple neurons present on each electrode. As such, we anticipate that these population  
926 recordings contain both single and multiunit activity.

927 For **Figure 4** and **Figure 4, Supplement 1** (Monkey J), each trial was defined as the 1200 ms following the reach  
928 target onset. Spike times were binned in 5 ms increments. For **Figure 5** (Monkey U) and **Figure 5, Supplement 1**  
929 (Monkey J), we aligned spike times to the go cue instead of target onset. To highlight oscillatory spiking activity, we  
930 defined each trial as the period occurring 400 ms prior to go cue and 100 ms after go cue. Spike times were binned  
931 in 2.5 ms increments for Monkey J and 5 ms increments for Monkey U; similar results were found for smaller bin  
932 sizes, and stronger smoothness regularization. In **Figure 5F** (Monkey U), we extended each trial duration to  $\pm 400$   
933 ms around the go cue, but otherwise kept the same parameters.

934 Tuning the regularization strength of on template smoothness ( $\lambda$ ) and warp magnitude ( $\mu$ ) was important to uncover  
935 the oscillations in premotor cortex. We used the cross-validation procedure described above to determine roughly  
936 appropriate values for these parameters; we increased the regularization strength further for the purposes of  
937 visualization and to be confident that the results were not due to overfitting.

## 938 Rat Motor Task

939 All procedures and experiments were reviewed and approved by the Harvard Institutional Animal Care and Use  
940 Committee. We analyzed data that was collected as part of a previously published study ([Dhawale et al. 2017](#)),  
941 which describes all experimental procedures and data collection protocols in greater detail.

942 Experimental subjects were female Long Evans rats, 3-8 months old at the start of the experiment (Charles River).  
943 Extracellular recordings were obtained from 16 chronically implanted tetrodes in the motor cortex. Signals were  
944 amplified and digitized on a customized head-stage, and sampled at 30 kHz. The head stage was attached to a  
945 custom-designed tethering system that allowed the animal to move freely within its cage. Before implantation, an  
946 automated behavioral training framework (described in [Poddar et al. 2013](#)) was used to train the rats on a timed  
947 lever-pressing task (described in [Kawai et al. 2015](#)) until asymptotic performance was achieved.

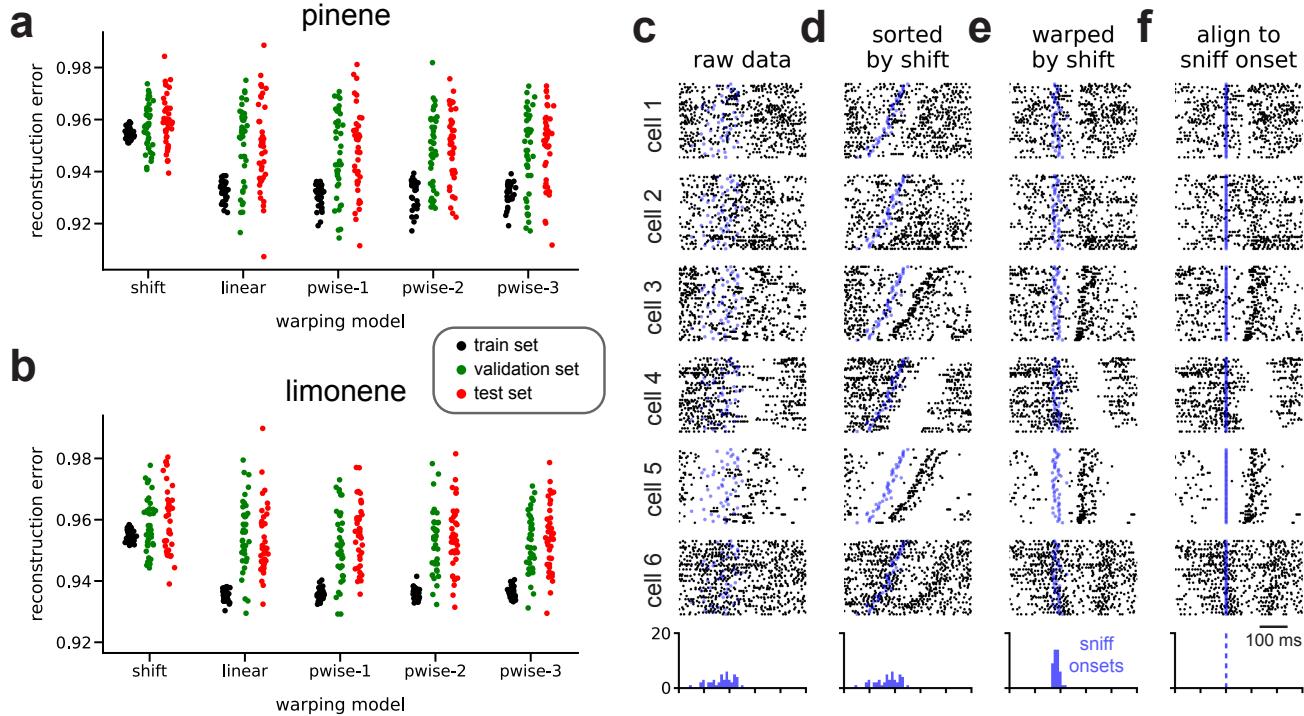
948 The tetrode drive was then surgically implanted and targeted to motor cortex, through a 4-5 mm diameter craniotomy  
949 made 2 mm anterior and 3 mm lateral to bregma. The tetrode array was lowered to a target depth of 1.85 mm. At  
950 the end of the experiments, the position of the electrodes was verified by standard histological methods—brains

951 were fixed via transcardial perfusion (4% paraformaldehyde in phosphate-buffered saline, Electron Microscopy  
952 Sciences) and the location of the electrodes was reconstructed by viewing mounted coronal sections (60 mm).

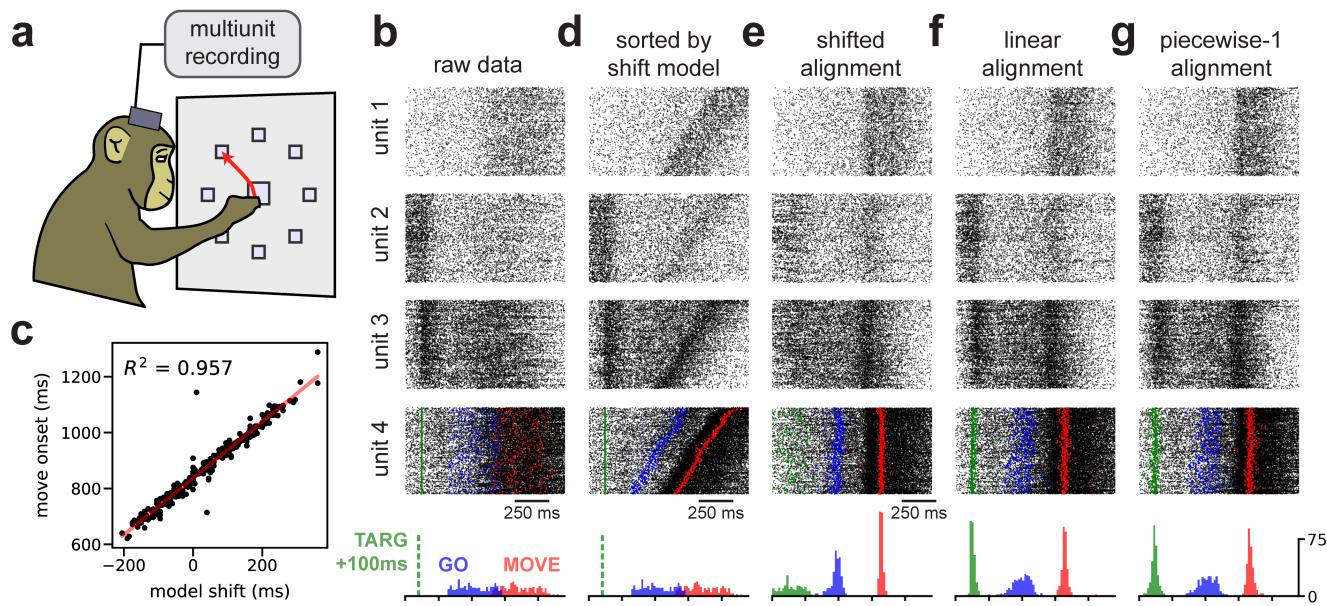
953 After 7 days of recovery post-surgery, training on the task resumed in the animal's home cage. Neural and behavioral  
954 data was recorded continuously during this time (12-16 weeks) with only brief interruptions (median time of 0.2 hr).  
955 Spikes were sorted using Fast Automated Spike Tracker (FAST), a custom algorithm designed for parsing long-term  
956 continuous neural recordings (for details, see [Dhawale et al. 2017](#)). We examined  $K = 1265$  trials, collected over a  
957 two day period.

958 Each trial was defined as the period starting 500 ms prior to the first lever press and 1500 ms after the first  
959 lever press. Spike times were binned in 10 ms increments for each unit. Raw spike counts were provided to  
960 the time warping algorithm; however, we observed similar results under various normalization schemes, such as  
961 soft-normalization ([Churchland et al. 2012](#)). All analyses of these data used a shift-only time warping model. The  
962 per-trial shift was constrained to be less than 10% of the total trial duration.

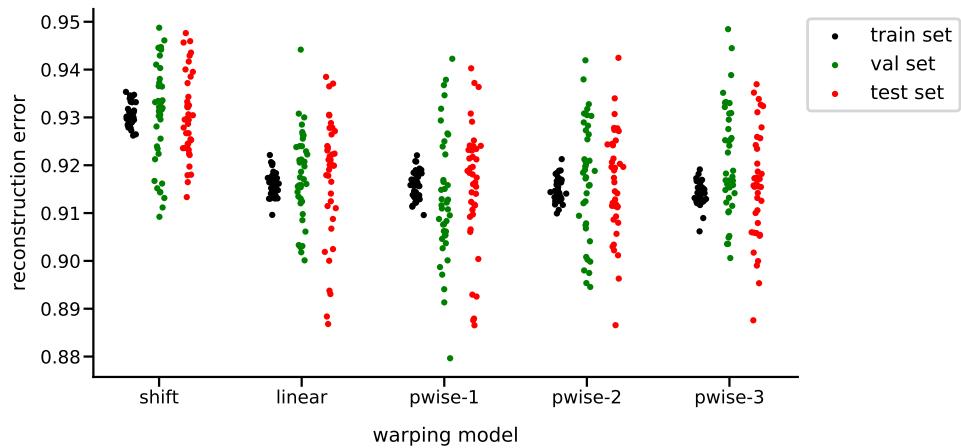
## 963 Supplemental Figures



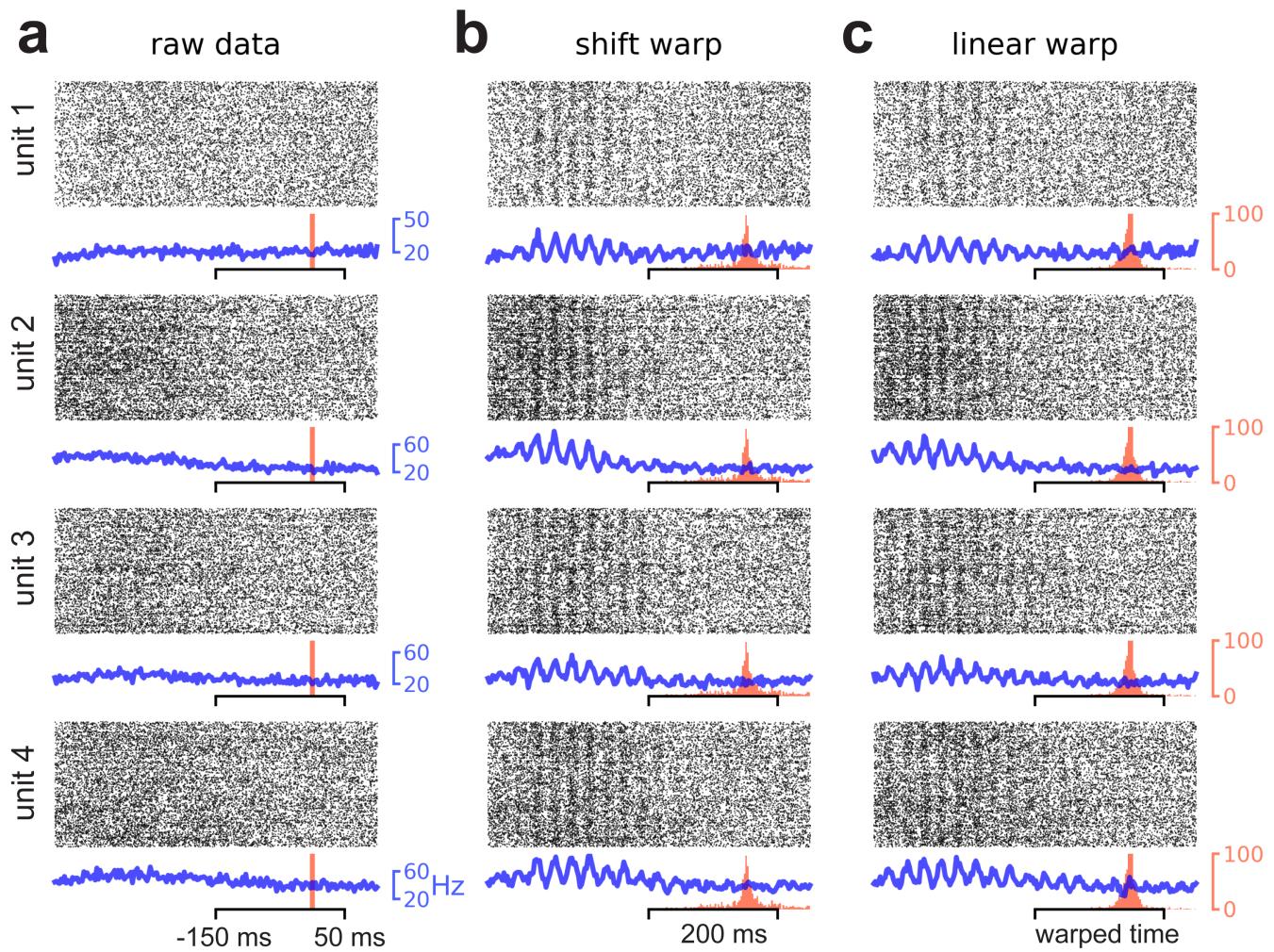
**Figure 3, Supplement 1. Nested cross-validation results and supplementary results on a different odorant.** (A) Nested cross-validation results for data presented in Fig. 2 (pinene,  $10^{-2}$  M). Vertical axis shows Euclidean norm of model residuals divided by norm of data for training and test sets. (B) Same as panel A, but computed on neural responses to a different odorant (limonene,  $10^{-2}$  M). (C-F) Same as Fig. 2B-E in the main text, but on  $K = 45$  presentations of limonene ( $10^{-2}$  M).



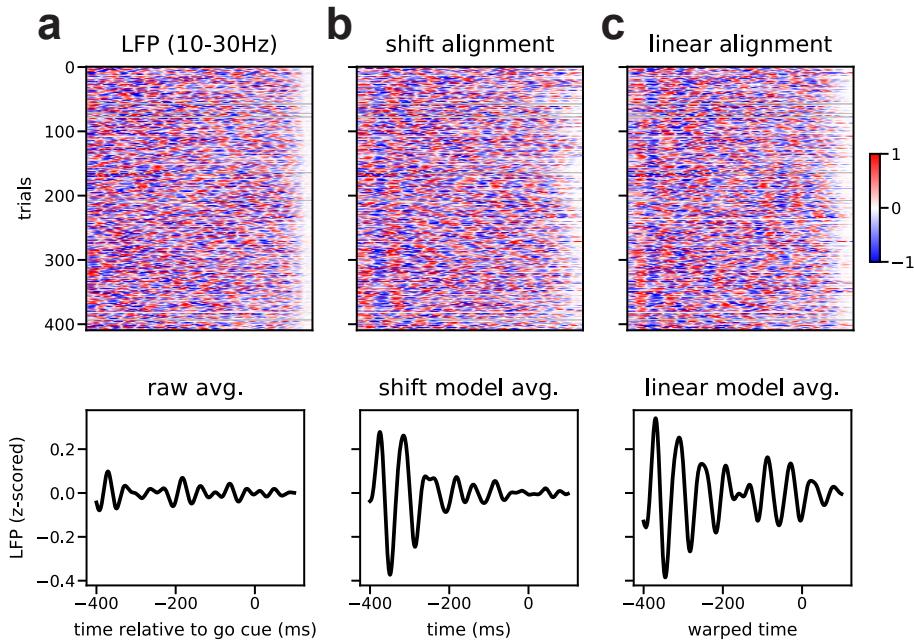
**Figure 4, Supplement 1. Replication of movement onset detection in primate reaching experiment.** All figure panels are directly analogous to Figure 4 in the main text and show comparable results. Instead of  $90^\circ$  reach trials (analyzed in the main text) only  $135^\circ$  reach trials were analyzed.



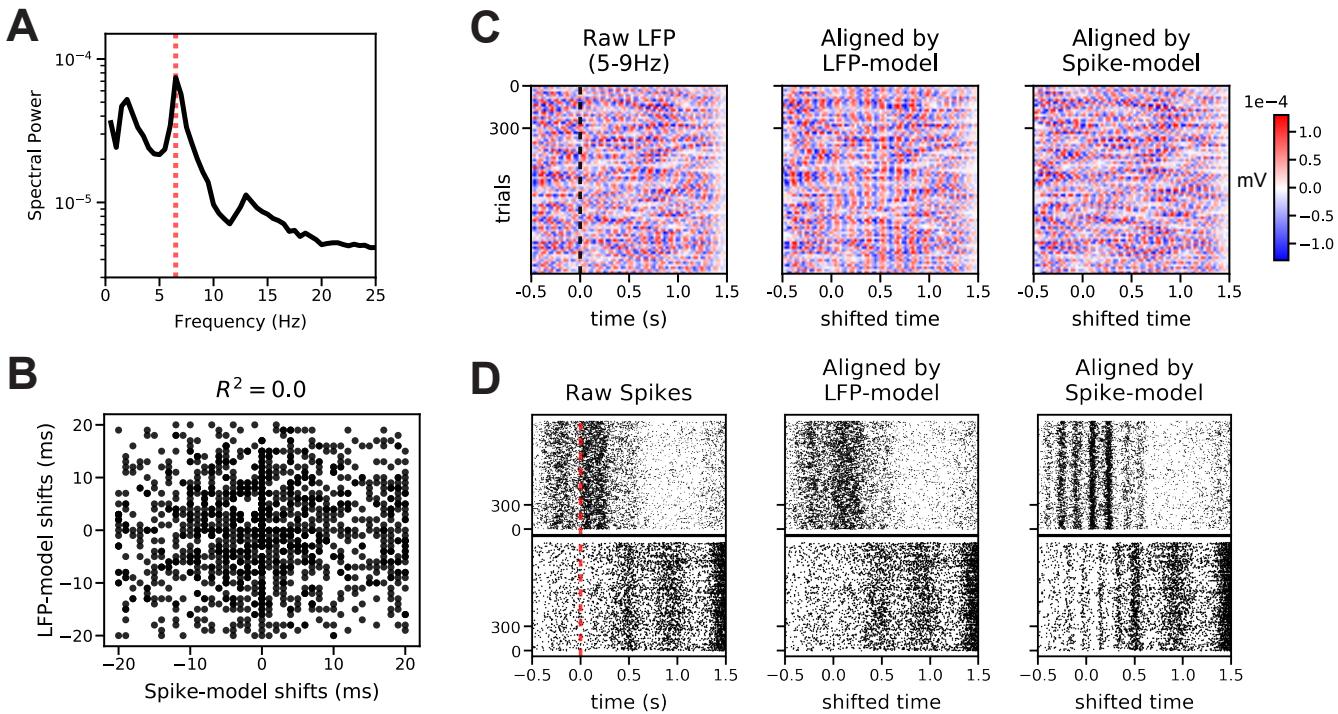
**Figure 4, Supplement 2. Nested cross-validation of primate reaching dynamics aligned to target onset.**



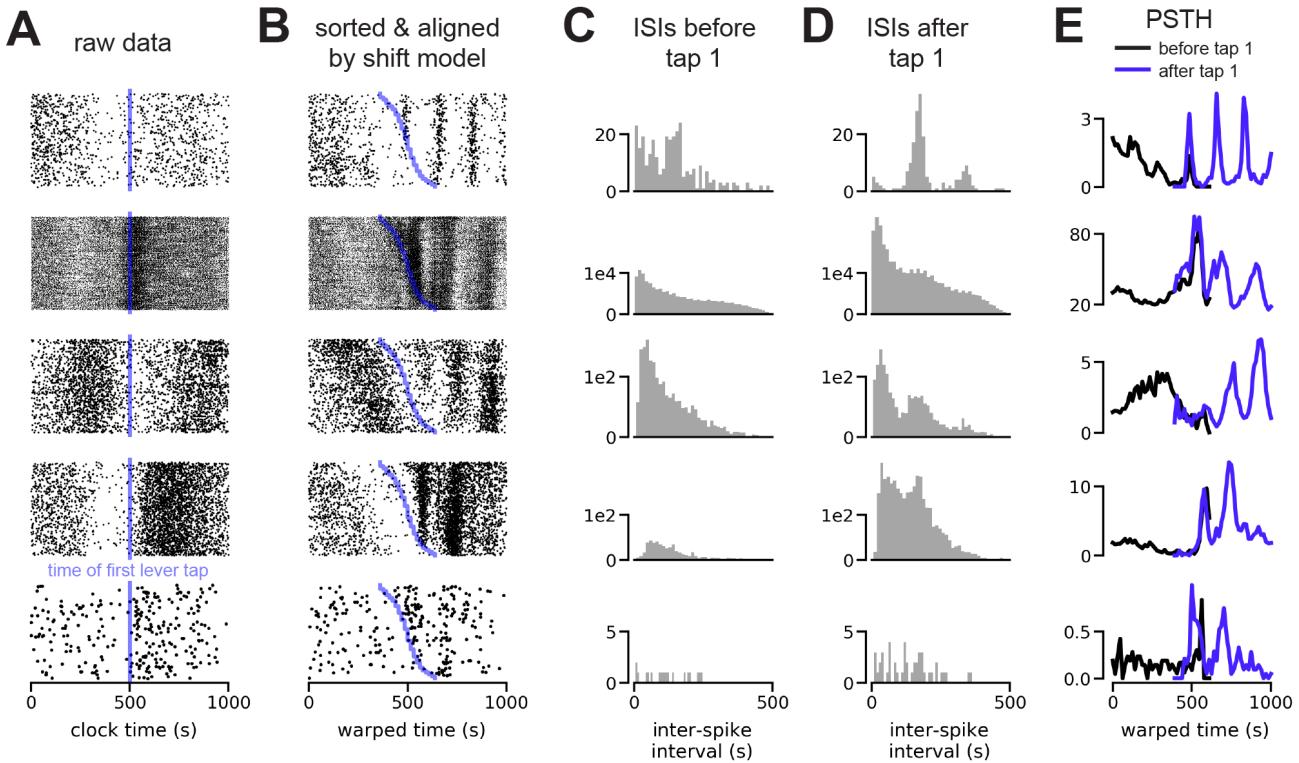
**Figure 5, Supplement 1. Oscillations in premotor cortex uncovered in data from a second nonhuman primate.** (A) Trial-by-time raster plots (black) and trial-averaged estimates of firing rate (blue) for four example multiunits. Red vertical line denotes the time that the go cue was delivered. (B-C) Same as panel A, except after spike times aligned by shift-only time warping (B) or linear time warping (C). Red histogram shows the distribution of go cue times after after the time warping transformation was applied. All four multiunits were held out during model fitting—the warping functions were fit to the remaining  $N = 95$  units and applied to the held out multiunit to generate the displayed raster plots.



**Figure 5, Supplement 2. Spike time oscillations in primate premotor cortex align with LFP.** (A) Top, LFP signal on all trials. The signal was obtained by averaging over all  $N = 96$  electrodes, z-scoring the signal within each trial, and then bandpass filtering (10-30 Hz; fifth-order Butterworth digital filter). Bottom, average LFP signal across trials. (B-C) Same as panel A, except after applying temporal alignments from a shift-only warping model (B) and a linear warping model (C). In both cases, time warping uncovered strong oscillations at  $\sim 18$  Hz—the same frequency of spike-level oscillations identified in Fig 4. Importantly, the warping models were fit *only* to binned spike times, demonstrating that the model generalized well to new data stream with fundamentally distinct features. This suggests that the spike-level oscillations described in Fig 4 are time-locked with LFP oscillations, in agreement with prior work (Murthy and Fetz 1992; Sanes and Donoghue 1993; Reimer and Hatsopoulos 2010; Pandarinath et al. 2018). All data were taken from the same animal subject shown in Fig 4.



**Figure 6, Supplement 1. LFP does not correlate with spike-level oscillations in rat motor cortex.** The LFP signal was highly correlated across all electrodes and thus averaged across electrodes before analysis. (A) Trial-averaged periodogram of the LFP signal. Dashed red line denotes 6.5Hz, illustrating a peak in the LFP spectrum that is similar to the frequency of spike-level oscillations in Fig 6. (B) One shift-only time warping model was fit to bandpassed-filtered LFP signals (LFP-model; fifth-order digital Butterworth, 5-9 Hz), and a second shift-only time warping model was fit to binned spike trains (Spike-model; same as Fig 6). The scatterplot demonstrates the per-trial shift parameters learned by these models were not correlated, suggesting that the spike-level oscillations are not phase-locked to LFP. (C) Bandpassed LFP as raw data (left; dashed line denotes first lever press), and same data aligned by LFP-model (middle) and Spike-model (right) time warping models. LFP is not reliably aligned by the time warping model fit to spiking data. (D) Raster plots from two example neurons (top and bottom rows), showing raw spike times (left; dashed line denotes first lever press), aligned by LFP-model (middle), and aligned by Spike-model (right). Spike-level oscillations are not revealed by the time warping model fit to LFP.



**Figure 6, Supplement 2. Five representative isolated units exhibiting stronger spike time oscillations following the first lever press.** (A) Raw spiking activity in a 1 second window around the first lever press. Vertical blue line denotes the time of the first lever press (manual alignment point). (B) Model-aligned spiking activity by shift-only warping, with trials sorted by the direction and magnitude of the learned shift. Blue line denotes the time of the first lever press on each trial. (C) Inter-spike interval (ISI) distributions during the 500 ms preceding the first lever press. (D) ISI distributions during the 500 ms following the first lever press. Note increased peak around ~150 ms, corresponding to increased oscillations at ~7 Hz. (E) Trial-averaged PSTHs for model-aligned spike times. Black lines denote PSTHs computed from spikes preceding the first lever press, while blue lines denote PSTHs computed from spikes following the first lever press. Note increased oscillatory dynamics following the lever press.