PNAS
www.pnas.org

Supplementary Information for

Temporal signals underlying a cognitive process in the dorsal premotor cortex

Román Rossi-Pool, Jerónimo Zizumbo, Manuel Alvarez, José Vergara, Antonio Zainos and Ranulfo Romo

Email: romanr@ifc.unam.mx (R.R.-P.); rromo@ifc.unam.mx (R.R.)

This PDF file includes:

Supplementary text
Figs. S1 to S7
Equations S1 to S17
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1820474116

Supplementary Information Text

Methods

Temporal Pattern Discrimination Task (TPDT)

The TPDT used here has been described before (1). Briefly, two monkeys (*Macaca mulatta*) were trained to report whether the temporal structure of two vibrotactile stimuli (P1 and P2) of equal mean frequency (5 Hz) were the same (P2 = P1) or different (P2 ≠ P1; Fig. 1A). Monkeys performed the task in blocks of trials in which the two stimulus patterns had a fixed mean frequency of 5Hz. The right arm, hand and fingers were held comfortably but firmly through the experiments. The left hand operated an immovable key (elbow at ~90°) and two push buttons in front of the animal, 25 cm away from the shoulder and at eye level. The centers of the buttons were located 7 and 10.5 cm to the left of the midsagittal plane. Stimuli were delivered to the skin of one digit from the distal segment of the right, restrained hand via a computer-controlled stimulator (2 mm round tip, BME Systems, Baltimore, MD). The initial skin indentation was 500μm (probe down event, "pd" in Fig. 1A). Vibrotactile stimuli consisted of trains of short mechanical pulses; each of those pulses consisted of a single-cycle sinusoid lasting 20ms. Time is always referenced to first stimulus onset (0 s corresponds to the start of P. In a trial, P1 and P2 were delivered consecutively to the glabrous skin of one fingertip, separated by a fixed inter-stimulus delay period of 2 s (from 1 to 3s in Fig. 1A). Each stimulus could be one of two possible patterns: a pattern of grouped pulses (upper trace of Fig. 1A; we will call it "G") or a pattern of extended pulses (lower trace of Fig. 1A, which will be "E"). Thus, in total there were four possible P1-P2 combinations, that we called classes: G-G (class 1, c1), G-E (class 2, c2), E-G (class 3, c3) and E-E (class 4, c4). These were presented in pseudo-random order to the monkeys across trials. The monkeys were asked to report whether P2 = P1 (match: combinations E-E and G-G) or P2 ≠ P1 (non-match: combinations E-G and G-E) after a fixed delay period of 2 s (from 4s to 6 s in Fig. 1A) between the end of P2 and the mechanical probe up from the skin (probe up event, "pu" in Fig. 1A). The "pu" was the go signal that triggered the animal's release of the key ("ku" in Fig. 1A). The monkey indicated their decision by pressing one of two push buttons with the left hand ("pb" in Fig. 1A,

lateral push button for P2 = P1, medial push button for P2≠P1). Because the two stimulus patterns had equal mean frequency over their full duration (1 s), the decision had to be based on comparison of their temporal structure. The animals were rewarded for correct decisions with a drop of liquid.

Light Control Task (LCT)

During this task, events proceeded exactly as described above and in Fig. 1A, except that when the probe touched the skin, one of the two push buttons was illuminated, indicating the correct choice (Fig. 5A). The light was turned off when the probe lifted off from the skin, triggering the hand/arm movement. The monkey was rewarded for pressing the previously illuminated button. Thus, stimuli and arm movements were identical to those in the active task but were cued by visual stimuli. The same mean stimulus frequency (5 Hz) used in the TPDT was delivered to the fingertip in the LCT.

Task Design and Performance

The TPDT is not a simple variation of the vibrotactile frequency discrimination task (VFDT, [2]). Some cognitive demands and the basic structure of the tasks are similar: both require attention to two separate vibrotactile stimuli (TPDT: P1, P2; VFDT: f1, f2), working memory and a comparison to reach the decision report. Nevertheless, the TPDT requires a very different transformation of the stimuli. Since they only differ by their temporal structure, any computation must be restricted to the internal structure to identify, categorize and discriminate the stimulus patterns (1). Additionally, the comparison process is significantly different between the two tasks. Whereas the VFDT can be solved by computing a difference between the parametric representation of the stimulus frequencies to indicate whether f1 > f2 or f1 <f 2, the TPDT offers no equivalent solution (in any trial P1 and P2 always have the same mean frequency). The TDPT demands a match (P2 = P1) vs. non-match decision (P2 ≠ P1). Thus, the comparison employs categorical representations of the stimulus patterns.

We quantified the average performance and its standard deviation across DPC recording sessions (85.8%). Even though each animal was trained between two and three years, this task was difficult enough not to allow 100% performance; this reflects the very high cognitive demands of the TPDT. To provide some context, the average training period to achieve similar performance levels for the VFDT was about six to eight months (2); for the vibrotactile detection task (3), the average time was two months. After training in the TPDT, the monkeys saturated their average performance around 85%; the performance was statistically the same for each class (1). Notably, repetition of the task during recording sessions did not improve performance. However, the performance for the LCT was consistently 100%; this reflects that this guided task was not cognitively demanding, which was part of the intended design. As a final observation, the animals were first trained in the LCT, and then gradually introduced to the TPDT. During the recording sessions in DPC (Fig. 1B), animals switched between performing the TPDT and the LCT.

Recordings

Neuronal recordings were obtained with an array of seven independent movable microelectrodes (2-3 MΩ; [2]) inserted into the DPC (Fig. 1B), contralateral (left hemisphere) or ipsilateral (right hemisphere) to the stimulated hand. We collected neuronal data in blocks of trials using different mean frequencies (1), but here for the analysis described below we will focus on the neuronal responses with the stimulus set illustrated in Fig. 1A (5 Hz); usually 20 trials per stimulus pair (c1; c2; c3; c4) per sesion. We remark that the recorded neuronal activity using block of trials (sets) with other mean frequencies (3, 6, 7, 10 and 15Hz) was collected after using both 5 Hz (n=1574) and light control sets (n=462). Hence, the number of neurons recorded under different mean frequencies were much fewer (n~200, [1]). Additionally, even fewer neurons were measured using at least three different mean frequency sets (n~50). Therefore, we restricted our temporal analysis to the neuron's responses recorded with the 5Hz stimulus set.

Recordings sites changed from session to session and the locations of the penetrations were used to construct surface maps of all the penetrations in DPC. This was done by

marking the edges of the small chamber (7 mm in diameter) placed above DPC. Recording of DPC neurons was done in the same manner in the TPDT and LCT.

Datasets

We recorded 1574 DPC neurons using the TPDT stimulus set with 5 Hz mean frequency. Additionally, we have a dataset of n=462 neurons that were tested in both the LCT and in TPDT using the 5Hz mean frequency set. These neurons were used to compare temporal and coding signals between the cognitive demanding TPDT and the non-demanding LCT.

For each neuron of the datasets (n=1574 and n=462), we calculated a time-dependent firing rate per trial using a 200 ms sliding window displaced every 50 ms, beginning 1 s before stimulus pattern P1 until the end of the trial (1.5 s after the push button press). It is important to keep in mind that each dataset is defined by four dimensions: N, number of neurons; C, stimulus conditions (classes, always 4); T, time (from -1 to 7.5s, always 170 bins); K, number of hit trials (for each class). Furthermore, we constructed a similar dataset with error trials for the 5 Hz TPDT stimulus set. Each recorded neuron had on average 2.5 error trials for a given class. A remarkable characteristic of this task design is the low number of stimulus condition (four classes), which were equally demanding for the subject. This design allowed us to have, on average, 17.5 hit trials (and 2.5 error trials) per stimulus class for each studied neuron.

Population Analysis

Single DPC neurons displayed a large repertoire of neuronal responses associated with one or several components of the TPDT (Fig. S1). Here, we focus our analysis in the temporal neuronal population signals. For each neuron, we averaged per class the time-dependent firing rate of hit trials (c1, c2, c3 or c4). Using the peri-stimulus time histogram (PSTH) of each neuron, we constructed pseudo-simultaneous population responses by combining neural data mostly recorded separately. For each time and class, the population response is defined by an N-dimensional vector in which each component represents the firing rate from a different neuron. This means that including all the recorded neurons (n=1574), we

obtained a 1574-dimensional firing rate vector that depended on the time and class ($\bar{r}(t,c)$). The population firing rates averaged over all hit trials ($\bar{r}(t)$) was an N-dimensional vector that measures the mean response for each neuron ($r^i(t)$) as a function of time. For the LCT, the control condition, the population response was a 462-dimensional firing rate vector.

In the next equations, we employed the following notation: $\bar{r}(t)$ is the firing rate average over all hit trials at each time bin (joining the four classes, c1-c4), $\bar{r}(t,c)$ is the firing rate average per class (grouping trials according to each of the four classes), $\bar{r}(t,P1)$ is the firing rate average per P1 stimuli (splitting trials according to P1-identity), $\bar{r}(t,P2)$ is the firing rate average per P2 stimuli (separating trials according to P2-identity) and $\bar{r}(t,Dec)$ is the firing rate average per decision (dividing trials according to decision). To further explain this measure, the i component, $r^i(t)$, represents the firing rate average of neuron i at time t, across all trials. Similarly, $r^i(t,Dec)$ represent the firing rate average of neuron i at time t across trials with the same decision outcome (P2=P1 or P2≠P1).

Instantaneous Coding Variances across the Population

At each time point, the population instantaneous coding variance (Var_{COD}, Figs. 1C and 4B, blue trace) was computed as the quadratic square sum of the firing rate fluctuations among classes and neurons:

$$Var_{COD}(t) = \frac{1}{N}\frac{1}{4}\sum_{i=1}^{N}\sum_{c=1}^{4}(r^i(t,c) - r^i(t))^2 \qquad \text{(Eq. S1)}$$

This metric, normalized per neuron, measures the population's variation of firing rate between classes at each time bin. With this definition, variation will be due to any class-related change in the population activity and also due to any residual noise.

To evaluate the influence of each kind of coding, we extended this metric to calculate the instantaneous variance associated to each specific task parameter. At each time bin, the population instantaneous P1 variance (Var_{P1}, Figs. 1C and 5B, cyan trace) was computed as the quadratic square sum of the firing rate fluctuations among P1 identity and neurons:

$$Var_{P1}(t) = \frac{1}{N}\frac{1}{2}\sum_{i=1}^{N}\sum_{P1=1}^{2}(r^i(t,P1) - r^i(t))^2 \qquad \text{(Eq. S2)}$$

Analogously, the population instantaneous P2 variance (Var_{P2}, Figs. 1C and 5B, light green trace) measures the firing rate fluctuations among P2 identity and neurons:

$$Var_{P2}(t) = \frac{1}{N}\frac{1}{2}\sum_{i=1}^{N}\sum_{P2=1}^{2}(r^i(t,P2) - r^i(t))^2 \quad \text{(Eq. S3)}$$

The population instantaneous decision variance (Var_{Dec}, Figs. 1C and 5B, black trace) measures the firing rate fluctuations among decision identity and neurons:

$$Var_{Dec}(t) = \frac{1}{N}\frac{1}{2}\sum_{i=1}^{N}\sum_{DEC=1}^{2}(r^i(t,Dec) - r^i(t))^2 \quad \text{(Eq. S4)}$$

The value of Var_{Cod} during the period immediately before P1 onset represented the inherent residual noise in the firing rate estimates (~2[sp/s]); to be interpreted as a degree of population coding, Var_{Cod} should be higher than this resting state variance (basal variance). The same reasoning applies to the other specific variances.

Instantaneous Temporal Variances across the Population

At each time point, the population instantaneous temporal variance (Var_{Temp}, Fig. 1D, Fig. 5C and Fig. 8A) with respect to the mean firing rate, was computed as the quadratic square sum between the mean firing rate for each time bin ($r^i(t)$) and the mean firing rate across the whole trial r^i (from -1s to 7.5s) among neurons:

$$Var_{Temp}(t) = \frac{1}{N}\sum_{i=1}^{N}\left(r^i(t) - r^i\right)^2 \quad \text{(Eq. S5)}$$

Additionally, the population instantaneous temporal fluctuation with respect to their basal response ($Fluc_{Temp}$, Figs. S2A and S7B, green trace) was computed as the quadratic square sum between the mean firing rate for each time bin ($r^i(t)$) and the mean basal firing rate r^i_{basal} (from -1s to 0s) among neurons:

$$Fluc_{Temp}(t) = \frac{1}{N}\sum_{i=1}^{N}\left(r^i(t) - r^i_{basal}\right)^2 \quad \text{(Eq. S6)}$$

$Fluc_{Temp}$ quantifies at each time point the temporal difference between the mean firing response and the basal activity, averaged across DPC neurons.

Except for the analysis shown in Fig. S2 and S7, in the remaining figures we computed the temporal variance of each neuron with respect to their mean firing rate (Eq. S5, Var_{Temp}).

Analogous results were found using r^i or r^i_{basal} as reference (compare Fig. 1 with Fig. S2 and Figs. 7 and 8 with Fig. S7).

Single Neuron Variances

The single neuron coding variance ($SNVar^i_{Cod}$) was calculated for each neuron *i* as the quadratic square sum over time of its firing rate fluctuations among classes:

$$SNVar^i_{Cod} = \frac{1}{T}\sum_{t=0}^{T}\sum_{c=1}^{4}(r^i(t,c) - r^i(t))^2 \qquad \text{(Eq. S7)}$$

For each individual neuron, $SNVar^i_{Cod}$ combines variance over classes and time bins. Additionally, we calculate the single neuron temporal fluctuation with respect to its mean firing rate ($SNVar^i_{Temp}$). This value computes for each neuron *i* the quadratic square sum over time between the mean firing rate for each time bin ($r^i(t)$) and the mean firing rate across the trial r^i (from -1s to 7.5s):

$$SNVar^i_{Temp} = \frac{1}{T}\sum_{t=0}^{T}(r^i(t) - r^i)^2 \qquad \text{(Eq. S8)}$$

$SNVar^i_{Temp}$ computes the temporal variance integrated over the 170-time bins from -1s to 7.5s. In Fig. S2 and S7 we calculated the single neuron temporal variance ($SNFluc^i_{Temp}$), with respect to the mean basal firing rate r^i_{basal} (from -1s to 0s):

$$SNFluc^i_{Temp} = \frac{1}{T}\sum_{t=0}^{T}(r^i(t) - r^i_{basal})^2 \qquad \text{(Eq. S9)}$$

Note the difference between these three metrics and those presented in the previous section. Instantaneous population variances average over neurons for each time bin; the metrics here (single neuron coding or temporal variances) average over time bins for each neuron.

Latency Variances

For each neuron, we calculated two different latency variances (Fig. 2): a latency of temporal variance (Lat_{Temp}) and a latency of coding variance (Lat_{Cod}). They correspond to the time at which the stimulus-driven temporal or coding variance (during P1, from 0s to 1s) becomes statistically different with respect to its basal values. Individual trials variance

distributions were generated at each time point using a time window of 200 ms sliding steps of 1 ms during P1 and were compared against the variance distributions obtained during the basal period (200 ms before P1 onset) using the receiver operating characteristic (ROC; [4]). The first-time window of at least 5 consecutive bins at which the area under the ROC curve (AUROC value) was significantly different from 0.5 (permutation test, $p < 0.05$) was considered as the latency variance. Furthermore, the time at which the temporal latency variance departs from its basal values coincides with the response latency (1). In addition, the coding latency variance is in concordance with the emergence of significant coding in individual neurons (1).

Principal Component Analysis (PCA)

The main aim of PCA is to find a new coordinate system in which the data can be represented in a more succinct and compact manner. In other words, the idea is to define a low-dimensional subspace that captures most of the variance of the high-dimensional neural space. A noteworthy feature of PCA analysis is its connection to Hebbian synaptic learning, which has been explored by other researchers (5, 6). Here, we provide an abbreviated technical description. To characterize how the population activity covaries across classes as function of time, we performed PCA (Figs. S4 and S6) over classes (c1, c2, c3 or c4) where we combined variance over classes and time (from -1s to 7.5s). PCA yields a new coordinate system for the N-dimensional data, in which the first coordinate accounts for as much of the variance of the neural population. The second coordinate accounts for as much of the remaining variance, and so on; however, each subsequent axis is restricted to be orthogonal to all previous axes. To obtain the new coordinate system, the covariance matrix of the N-dimensional data must be diagonalized. The firing rate covariance matrix summarizes the second-order statistics of the neuronal data. It was obtained averaging over time bins, *t*, and classes, *c*:

$$C_{ij} = \frac{1}{TC-1} \sum_{t=-1}^{t=7.5} \sum_{c=1}^{4} (r^i(t,c) - r^i(t))(r^j(t,c) - r^j(t)) \qquad \text{(Eq. S10)}$$

where *T* denotes the number of time bins and *C* is the number of classes (4 in our task), $r^i(t,c)$ denotes the trial-averaged firing rate of the neuron *i*, under class *c*, at time *t*, and

$r^i(t)$ is the firing rate average of neuron *i* across classes at time *t*. In Fig. S4, PCs were calculated for TPDT (n=1574); in Fig. S6, PCs were calculated for TPDT and LCT, with the subpopulation of neurons recorded in both (n=462). The diagonalization of the covariance matrix, $C=UDU^T$, yields a new coordinate system given by the columns of the matrix *U*. We refer to the columns of *U* as the principal components (PCs). On the other hand, *D* is a diagonal matrix of positive values. The diagonal elements of *D* give the amounts of population activity variance captured by the corresponding PCs. We then ordered the PCs depending on this amount of variance captured. The projection of the N-dimensional data onto the *k*th PC is given by:

$$PC_k(t,c) = \sum_{i=1}^{N} U_k^i . (r^i(t,c) - r^i(t)) \qquad \text{(Eq. S11)}$$

where U_k^i is the *i* element of the *k* PC (U_k). Therefore, the PCs are linear readouts of the population activity; in other words, they are linear combinations of the firing rates of the individual neurons. Thus, the contribution of each neuron to a given PC_k is given by the *i*th element of U_k. These PCs can be thought of as a low-dimensional description of the population activity in this coding subspace.

Temporal Demixed Principal Component Analysis (dPCA)

The algorithmic details and mathematical justification for this method were outlined previously (7, 8). The method has a supervised and unsupervised part. In brief, dPCA decomposes the neural activity by different chosen task variables to compute marginalized covariance matrices (this is the supervised part, similar to choosing the variables to fit a linear model). Afterwards, it carries out a principal component-like analysis over those matrices (this is completely unsupervised). In this work, we only marginalized the population activity ($\bar{X}$) with respect to time ($\overline{X_t}$). To calculate the marginalization average, we use the N-dimensional population activity:

$$\overline{X_t} = \bar{r}(t) - \bar{r} \qquad \text{(Eq. S12)}$$

$\bar{r}$ is a vector with the mean firing rate over the whole task (from -1s to 7.5s), $\bar{r}(t)$ is the firing rate average over all hit trials at each time bin (joining the four classes, c1-c4). Thus, the *i* component, $r^i(t)$, represents the firing rate average of neuron *i* at time *t*, across all

trials. Once time marginalization is performed, dPCA finds separate decoder and encoder matrices for $\overline{X_t}$, by minimizing with reduced-ranked regression the term:

$$L_t = \|\overline{X_t} - \mathrm{F}_t \mathrm{D}_t \bar{X}\|^2 \qquad \text{(Eq. S13)}$$

where $\bar{X}$ is the centered whole population data matrix (i.e. the average activity of each neuron is 0). The solution of this problem can be obtained analytically in terms of singular value decompositions (7). Each temporal component of D_t can be ordered by the amount of explained variance. The most prominent decoding axis is called the 1st demixed temporal principal component (1st temp-dPC). To avoid dPCA overfitting, we introduced a regularization term and performed cross-validation to choose its value.

To obtain Figs. 4A and 7D, we projected the N-dimensional data for a given class ($\bar{r}(t,c)$) onto the five most prominent temporal decoding axes (k). These projections were given by:

$$dPC_k^{temp}(t,c) = \sum_{i=1}^{N} D_{k,temp}^i (r^i(t,c) - r^i(t)) \qquad \text{(Eq. S14)}$$

where $dPC_{k,temp}^i$ is the *i-th* component of the *k-th* most relevant temporal demixed axis. In Figs. 4A and 7D we used dPCA to identify the population temporal signals that explain most variance of the whole task time-marginalized covariance matrix ($\overline{X_t}$).

Trial Variability in Temporal Components

To further study the single trial variability on each temporal decoding axis we separated the data into training and testing sets (7). On each iteration, we randomly chose one trial from each neuron for each class, defining a single trial population activity X_{test}. The remaining trials were averaged to form a training population activity $\overline{X_{train}}$. We constructed training matrices with temporal marginalized population activities (Fig. 8C). Note that X_{test} and $\overline{X_{train}}$ have the same dimensions (*Nx(TxC), C=4 classes*).

Next, we calculate temporal dPCA on the corresponding $\overline{X_{train}}$ to obtain the different decoding axes. We then projected X_{test} onto these temporal axes. We repeated this procedure 1000 times. With all the iterations we obtained the mean value (wider line) and standard deviation (SD, colored shading) at each time bin for each class. On Fig. 8C, we displayed the mean and SD of all the projections.

Error Trial Variability in Temporal Components

We extended the analysis of single hit trial variability to error trials. We remark that each recorded neuron had on average 2.5 error trials for a given class. In this case, we used the temporal axes calculated with dPCA using the hit trials' whole activity ($\bar{X}$). Analogously, we now randomly chose one error trial per iteration for each neuron in each class. This built a single error trial population activity X_{error}. In contrast to the previous procedure (for hits), $\overline{X_{train}}$ is always just $\bar{X}$. The rest was done in the same manner. On Fig. 8D, we displayed the mean and SD of all the error trial projections.

In Fig. 8A-B and S7A-B we show Var_{Temp} (Eq. S5) and $SNVar^{i}_{Temp}$ (Eq. S8) for errors; the procedure to calculate them is same as described before. In this panels, we used the n=547 neurons with at least 3 errors in each class (12 errors in total).

Explained Variance, Noise, and Dimensionality

The fraction of explained variance for each temporal decoder (F_t, Eq. S13, [7]) is:

$$R^2 = \frac{\|\bar{X}\|^2 - \|\bar{X} - F_t D_t \bar{X}\|^2}{\|\bar{X}\|^2} \quad \text{(Eq. S15)}$$

We used this equation to compute the fraction of variance explained by each temporal dPC. Along the text, we refer to this fraction as the explained total variance (ETV); these are the percentages reported in Figs. 4, 7 and S6. It should be noted that $\bar{X}$ includes the activity of the whole task.

Further, there are two other matrices on which we perform variance calculations: $\check{X}$ and X_{noise}. They are related to each other, and to the $\bar{X}$ described previously, by:

$$\check{X} = \bar{X} + X_{noise} \quad \text{(Eq. S16)}$$

The full neural activity matrix $\check{X}$ contains *K=14* hit trials per class and neuron (n=1574). The value of *K* was selected to be the minimum number of trials present for all neurons across all classes. Thus, dimensions of $\check{X}$ are *Nx(TxCxK)*. Note that $\bar{X}$ in Eq. S16, is the same as described above but with their *NxCxT* unique values repeated by the number of trials *K* (7). For neurons that were recorded on the same session, we attempted to select the

same *K* trials for all of them. However, we obtained similar results when we did a completely random selection of the *K* trials.

In order to estimate the total variance (PSTH + noise variance) explained by each temporal decoding axis (ESNV: explained signal + noise variance), it was necessary to replace $\bar{X}$ by $\check{X}$ in Eq. S15. These values are plotted for each component (dPCs or PCs) in Fig. 4B (TPDT), 7B (LCT) and S6 (TPDT and LCT). The cumulative fraction of total variance explained by the first q components was calculated using the same formula.

The percentage of variance related to trial variability could be estimated as: $(\|X_{noise}\|^2 / \|\check{X}\|^2)$. The red dotted lines in Figs. 4B and 7B were obtained as its complement: $(1 - \|X_{noise}\|^2 / \|\check{X}\|^2)$; they represent an estimate of the task related variance. Additionally, we estimated the percentage of noise variance explained by each decoding axis using a variation of Eq. S15:

$$R^2_{noise} = \frac{\|X_{noise}\|^2 - \|X_{noise} - FDX_{noise}\|^2}{\|\check{X}\|^2} \qquad \text{(Eq. S17)}$$

This noise projection variance used the same decoder components as the signal + noise calculation. The value of R^2_{noise} for the projections of the first decoding PCs are shown in Figs. 3B (TPDT, grey line), 6B (LCT) and S6 (TPDT and LCT). Using this measure, we estimated the percentage of noise variance explained by each decoder. In the TPDT, 12 components are at least 50% above this value. On the other hand, in the LCT, only 6 components are above. This gave us an estimate of the dimensions of the temporal and coding dynamics (9).

Given the experimental protocol, these noise variance calculations are the best approach to estimate its real value. However, the variability that we attributed to "noise" could have been associated with other processes, which are not the scope of this study (10).

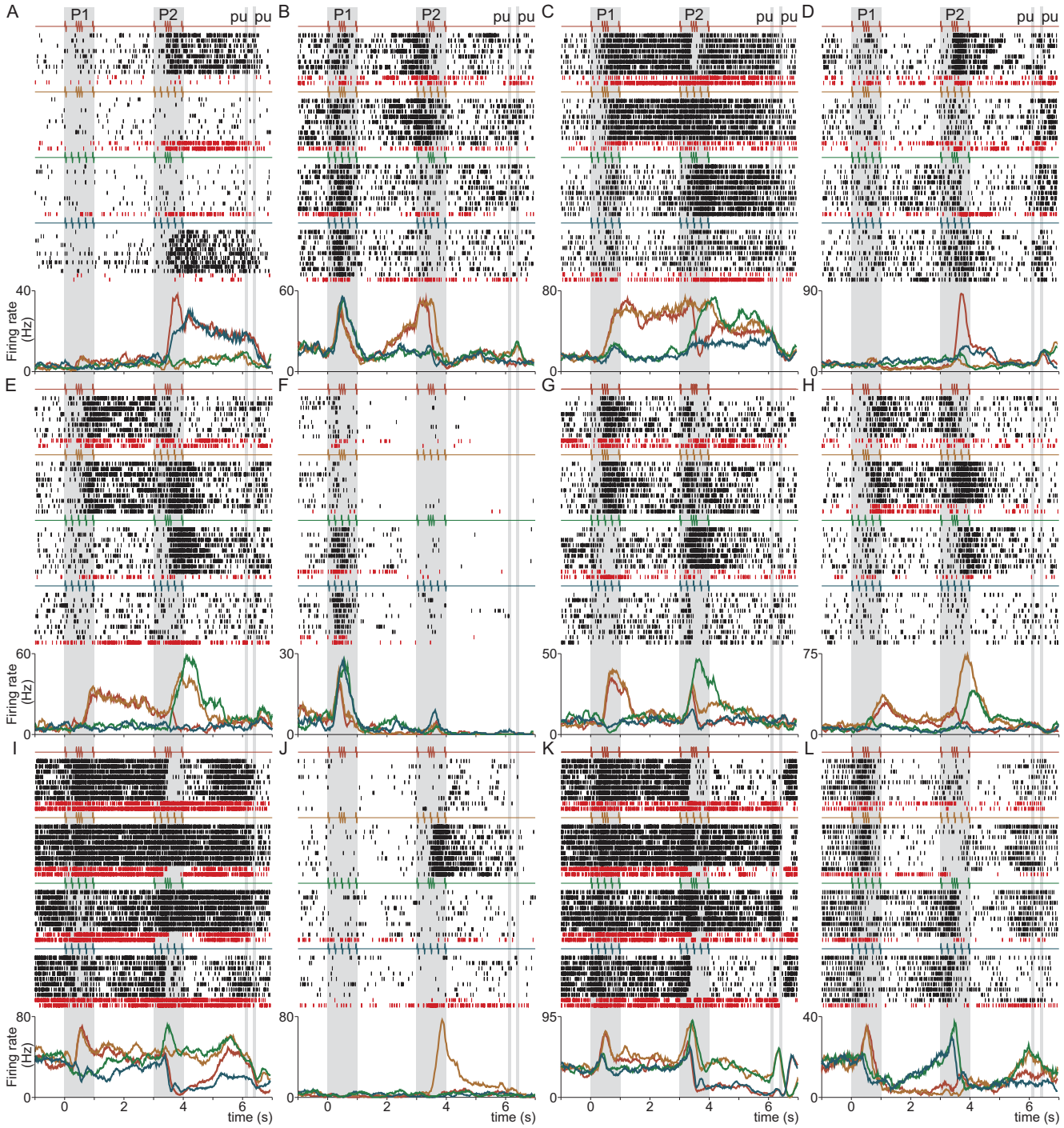

Fig. S1. Heterogeneity of Single DPC Neuron Parameter Coding during the TPDT. Related to Fig. 1. Neurons display heterogenous coding and mixed selectivity of different task parameters. (A-L) Raster plots of twelve exemplary neurons sorted according to the four possible combinations of G and E stimulus patterns delivered during P1 and P2. Each row of ticks is one trial, and each tick is an action potential. Trials were randomly delivered and were sorted by class afterward (only 10 out of 20 trials per class are shown). Correct and incorrect trials are indicated by black and dark red ticks, respectively. Traces below the raster plots are average per-class firing rates (PSTHs) for each neuron. Each color refers to one of the four possible stimulus combinations of G and E; the resulting four classes are c1 (G-G, red), c2 (G-E, orange), c3 (E-G, green) and c4 (E-E, blue).

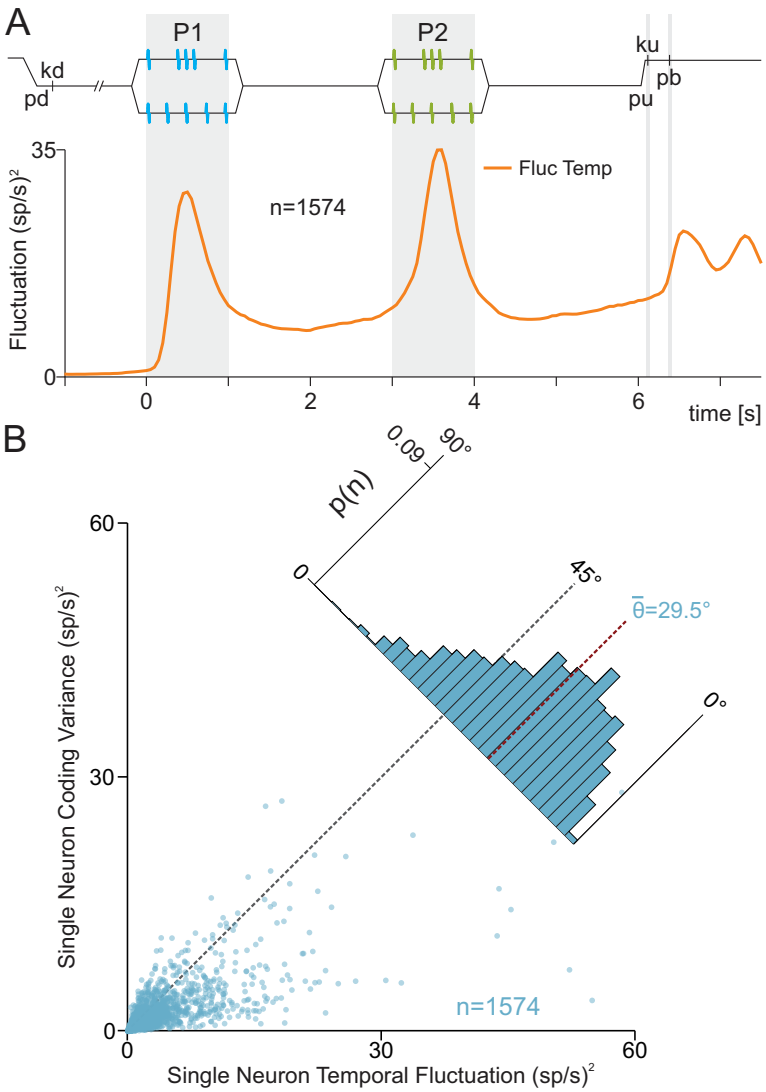

Fig. S2. Single DPC Neurons and Population Temporal Variance. Related to Fig. 1. (A) Population instantaneous temporal variance ($Fluc_{Temp}$, red trace, Eq. S6), computed as the quadratic square sum between the mean firing rate at each time bin ($r^i(t)$) and the mean basal firing rate: r^i_{basal} (from -1s to 0s). (B) Single neuron temporal variance ($SNFluc^i_{Temp}$, Eq. S9, x-axis) is plotted against single neuron coding variance (Eq. S7, y-axis, $SNVar^i_{Cod}$). Each dot corresponds to one neuron (n=1574). Diagonal line (45°) indicates equality between both types of variances. Inset histogram shows angular distributions for the population (<θ>=29.5º). In general, single neuron temporal variance is higher than coding variance. Similar results were found computing the temporal variance with respect to the mean firing rate across the whole task (Eq. S5, Var_{Temp}; Eq. S8, $SNVar^i_{Temp}$).

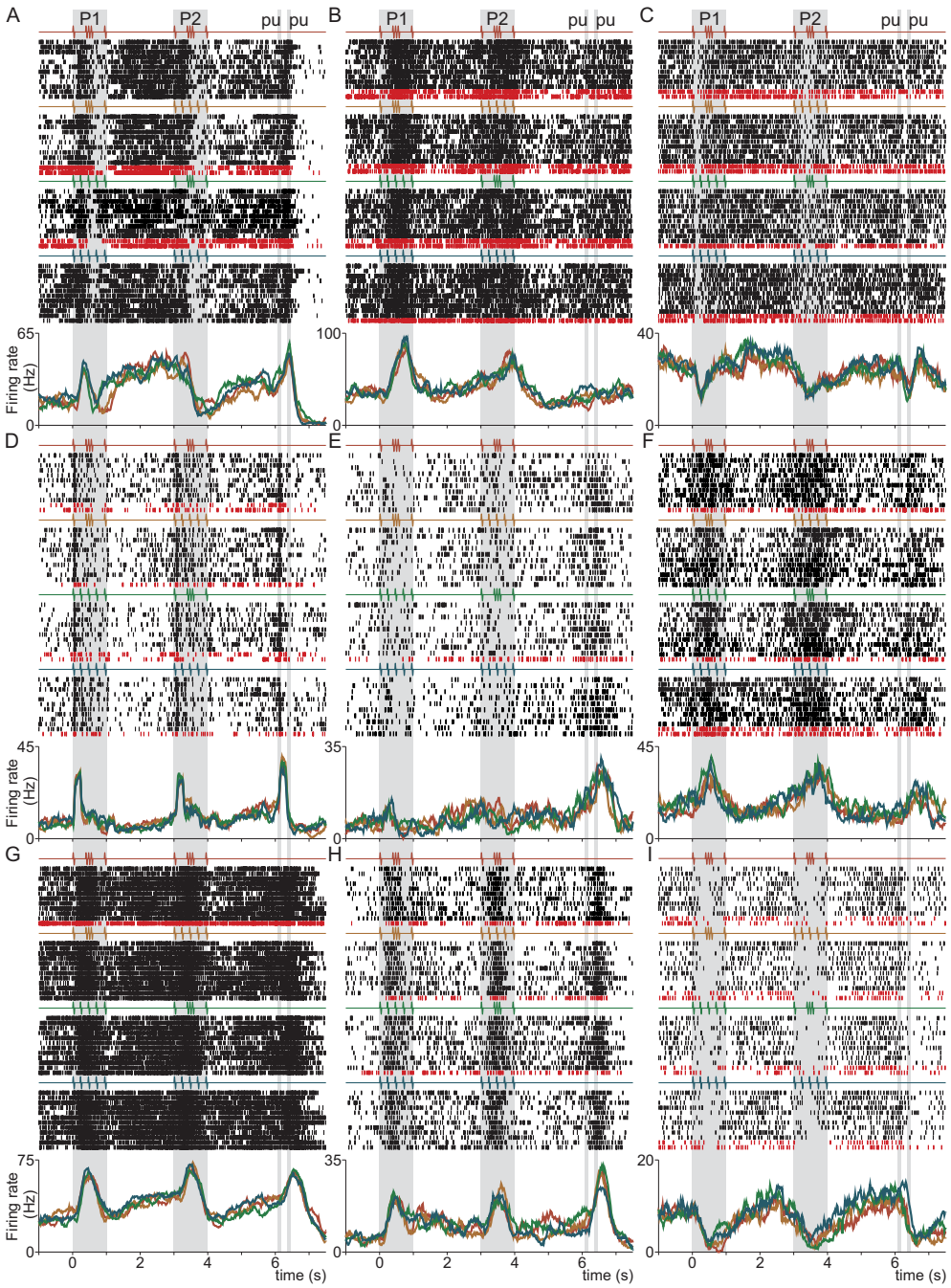

Fig. S3. Single DPC Neurons with Notable Temporal Activity in TPDT. Related to Fig. 3. (A-I) Raster plots of nine additional neurons with exemplary temporal dynamics. Trials were sorted according to the four possible combinations of G and E stimulus patterns delivered during P1 and P2. Each row of ticks is one trial, and each tick is an action potential. Trials were randomly delivered and were sorted by class afterward (only 10 out of 20 trials per class are shown). Correct and incorrect trials are indicated by black and dark red ticks, respectively. Traces below the raster plots are average per-class firing rates (PSTHs) for each neuron. Each color refers to one of the four possible stimulus combinations of G and E; the resulting four classes are c1 (G-G, red), c2 (G-E, orange), c3 (E-G, green) and c4 (E-E, blue). Neurons display temporal dynamics without coding task parameter.

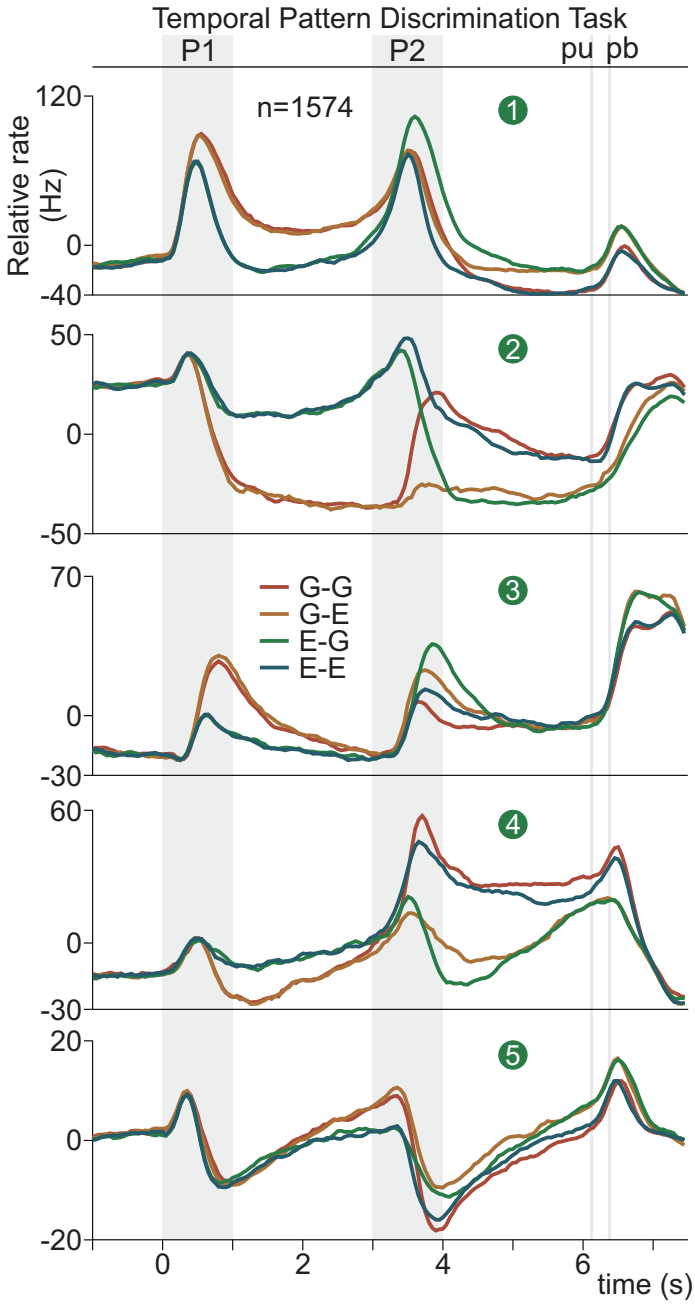

Fig. S4. Principal Components calculated from the DPC neuronal population during the TPDT. Related to Fig. 4. Principal component analysis (PCA) was applied to the temporal marginalized covariance matrix obtained from the whole TPDT (from -1 to 7.5s). The population activity, sorted by class identity, was projected onto each principal axis. Components were ordered according to their explained total variance [ETV, Eq. S15]. First five principal components (PCs): 1st PC, ETV 24.1%; 2nd PC, ETV 13.9%; 3rd PC, ETV 11.4%; 4th PC, ETV 10.2%; 5th PC, ETV 9.5.

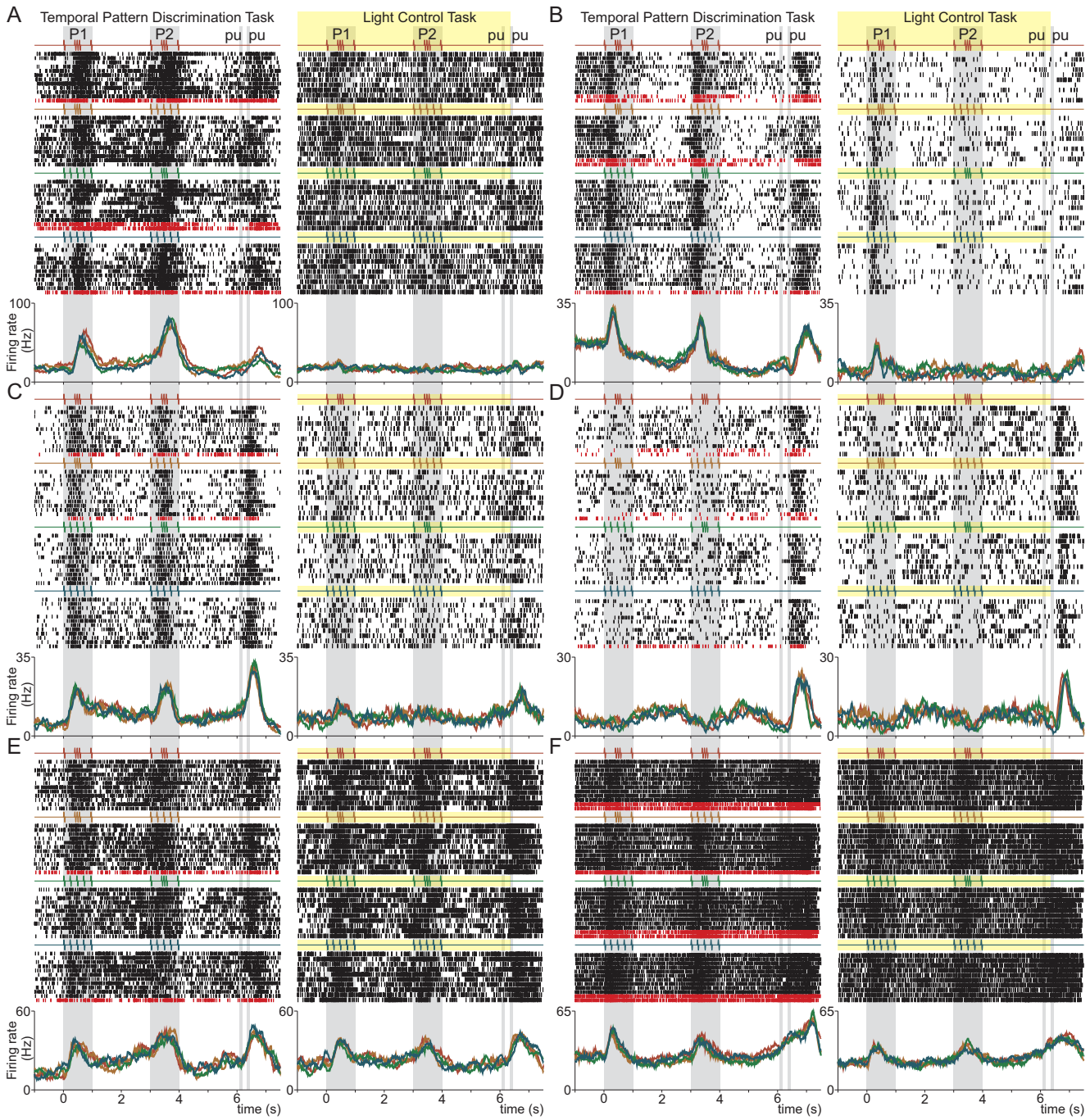

Fig. S5. Single DPC Neurons with Notable Temporal Activity compared between TPDT and LCT. Related to Fig. 6. (A-F) Raster plots of six additional exemplary neurons with temporal dynamics tested in both conditions: TPDT (left panels) and LCT (right panels). Responses are sorted according to the four possible combinations of G and E stimulus patterns delivered during P1 and P2. Correct and incorrect (only in TPDT) trials are indicated by black and dark red ticks, respectively. Traces below the raster plots are average per-class firing rates (PSTHs) for each neuron. Each color refers to one of the four possible stimulus combinations of G and E; the resulting four classes are c1 (G-G, red), c2 (G-E, orange), c3 (E-G, green) and c4 (E-E, blue).

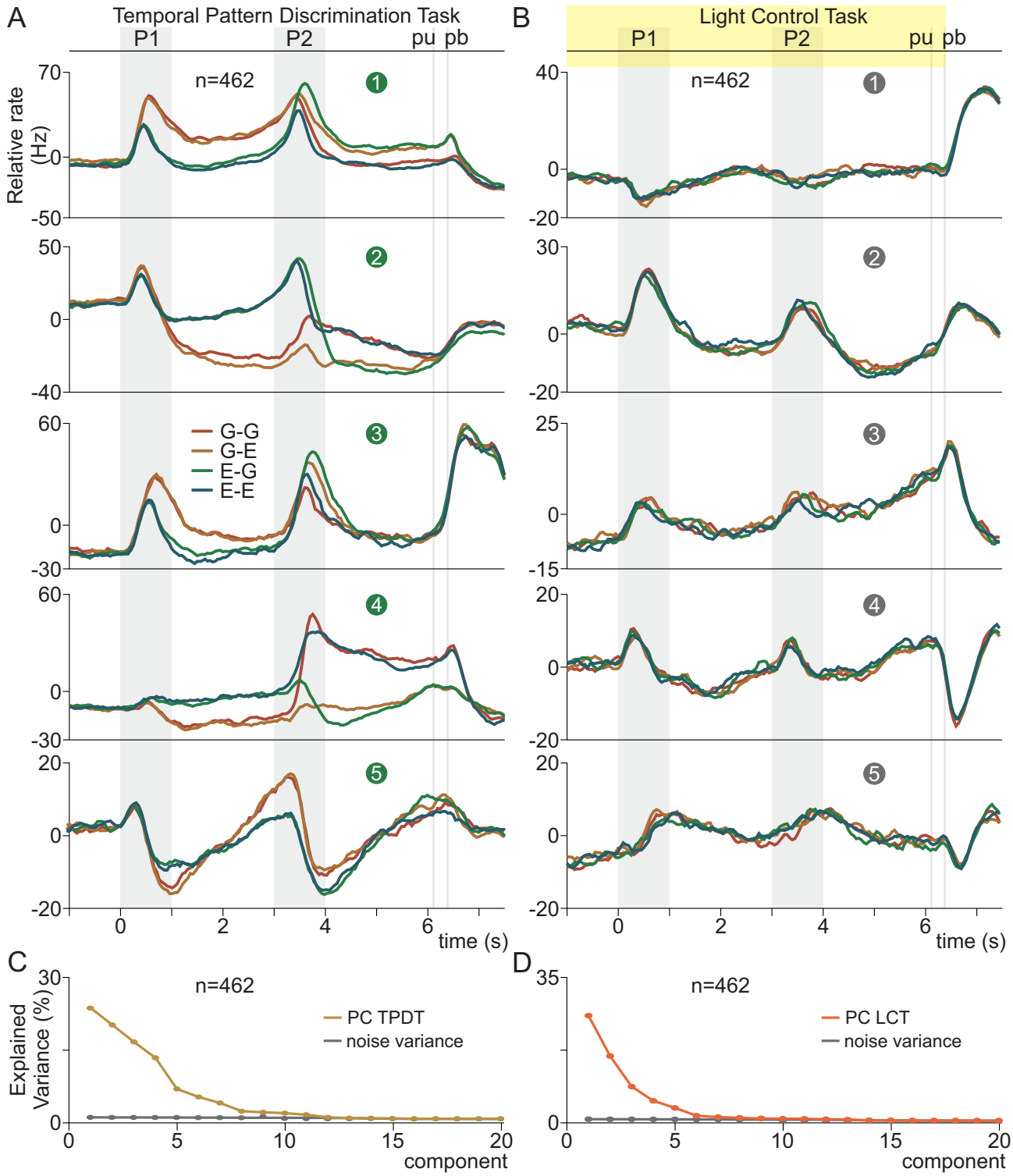

Fig. S6. Principal Components calculated from the DPC neuronal population during the TPDT and LCT. Related to Fig. 7. In this figure we restricted our analysis to the population of n=462 neurons that were recorded both in the temporal pattern discrimination task (TPDT) and light control task (LCT). Principal component analysis (PCA) was applied to the covariance matrices obtained for the whole duration (from -1 to 7.5s), for both tasks (TPDT and LCT). The population activity, sorted by class identity, was projected onto each principal axis. Components were ordered according to their explained total variance [ETV, Eq. S15]. (A) TPDT (n=462). First five principal components (PCs): 1st PC, ETV 22.8%; 2nd PC, ETV 14.3%; 3rd PC, ETV 10.9%; 4th PC, ETV 12.1%; 5th PC, ETV 9.8. (B) LCT (n=462). First five principal components (PCs): 1st PC, ETV 30.7%; 2nd PC, ETV 19.6%; 3rd PC, ETV 10.8%; 4th PC, ETV 6.5%; 5th PC, ETV 4.3%. (C-D) Show signal + noise variance (ESNV) for principal components during TPDT (golden) or LCT (orange) using the same group of neurons (n=462). Grey trace represents the percentage of noise variance explained by each PCs (Eq. S17). The number of significant principal components above noise was 11 during TPDT and 6 during LCT. The dimension of the network's dynamic increased during task demand.

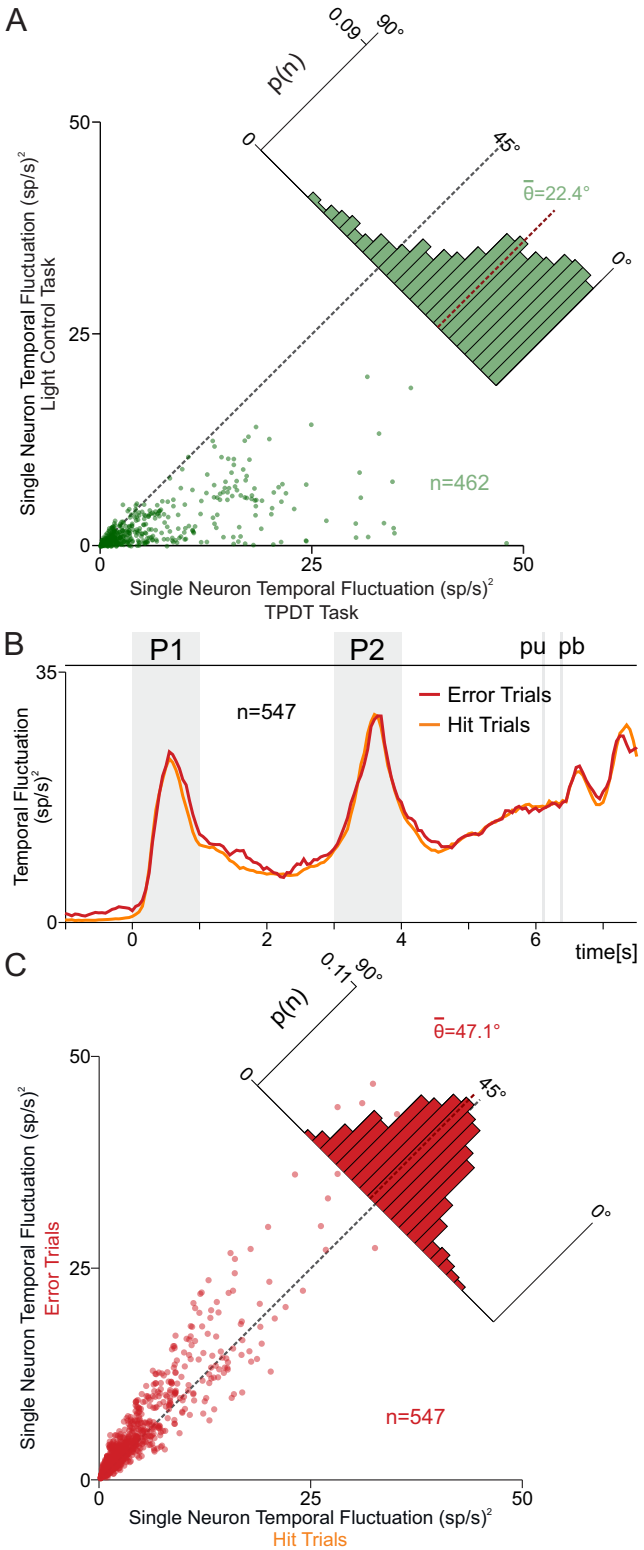

Fig. S7. Temporal Fluctuations during LCT and error trials. Related to Figs. 7 and 8. (A) In this panel we restricted our analysis to the population of n=462 neurons that were recorded both in the temporal pattern discrimination task (TPDT) and light control task (LCT). Single neuron TPDT temporal fluctuation ($SNFluc^{i}_{Temp}$, Eq. 9, x-axis) is compared against single neuron LCT temporal fluctuation ($SNFluc^{i}_{Temp}$, Eq. 9, y-axis). Each dot corresponds to one neuron tested in both conditions (n=462). Diagonal line (45°) indicates equality between both condition variances. Inset histogram show angular distributions for the population (<θ>=22.4º). $SNFluc^{i}_{Temp}$ is much higher in TPDT than LCT. (B-C) In these panels we restricted our analysis to the population of n=547 neurons with at least 3 error trials for each of the four classes. (B) Population instantaneous temporal fluctuation for hit and error trials ($Fluc_{Temp}$, red and orange trace respectively, Eq. 6), computed as the quadratic square sum between the mean firing rate at each time ($r^{i}(t)$) in hit and error trials and the mean basal firing rate (r^{i}_{basal}). (C) Single neuron TPDT temporal fluctuation ($SNFluc^{i}_{Temp}$, Eq. 9) during hit trials (x-axis) is compared against error trials (y-axis). Each dot corresponds to $SNFluc^{i}_{Temp}$ of one neuron computed in both conditions (n=547). Diagonal line (45°) indicates equality between hit and error variances. Inset histogram show angular distributions for the population (<θ>=47.1º). $SNFluc^{i}_{Temp}$ distributions are statistically the same in hit and error trials.

References

1. Rossi-Pool R, et al. (2016) Emergence of an abstract categorical code enabling the discrimination of temporally structured tactile stimuli. *Proc Natl Acad Sci USA* 113:E7966–E7975.

2. Romo R, Brody CD, Hernández A, Lemus L (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399:470-473.

3. de Lafuente V, Romo R (2006) Neural correlate of subjective sensory experience gradually builds up across cortical areas. *Proc Natl Acad Sci USA* 103:14266-14271.

4. Green, DM, Swets, JA (1966) Signal Detection Theory and Psychophysics (Wiley, New York).

5. Oja E (1992) Principal components, minor components, and linear neural networks. *Neural Network* 5:927–935.

6. Diamantaras, KI, Kung SY (1996) Principal Component Neural Networks: Theory and Applications (Wiley, New York).

7. Kobak D, et al. (2016) Demixed principal component analysis of neural population data. *eLife* 5: e10989.

8. Rossi-Pool R, et al. (2017) Decoding a decision process in the neuronal population of dorsal premotor complex. *Neuron* 96:1432–1446.

9. Machens CK, Romo R, and Brody CD (2010) Functional, but not anatomical, separation of ‘‘what’’ and ‘‘when’’ in prefrontal cortex. *J Neurosci* 30:350–360.

10. Carnevale F, de Lafuente V, Romo R, Parga N (2012) Internal signal correlates neural populations and biases perceptual decision reports. *Proc Natl Acad Sci USA* 109:18938-18943.