Demixed PCA, or how to tie dimension reduction to experimental conditions

Mihály Bányai
Population Activity Study Group
MTA KOKI, 04.15. 2016.

Demixed Principal Component Analysis

Wieland Bren
Ecole Normale Supérieure,
Champalimaud Neuroscien
Lisbon, Portug

Ec
Champalim

Demixed principal component analysis of population activity in higher cortical areas reveals independent representation of task parameters

Dmitry Kobak[1,*], Wieland Brendel[1,2,*], Christos Constantinidis[3], Claudia E. Feierstein[1], Adam Kepecs[4], Zachary F. Mainen[1], Ranulfo Romo[5], Xue-Lian Qi[3], Naoshige Uchida[6], and Christian K. Machens[1]

[1]Champalimaud Centre for the Unknown, Lisbon, Portugal
[2]École Normale Supérieure, Paris, France
[3]Wake Forest University School of Medicine, Winston-Salem, NC, USA
[4]Cold Spring Harbor Laboratory, NY, USA
[5]Universidad Nacional Autónoma de México, Mexico
[6]Harvard University, Cambridge, MA, USA

[*]Equal contribution

October 2014

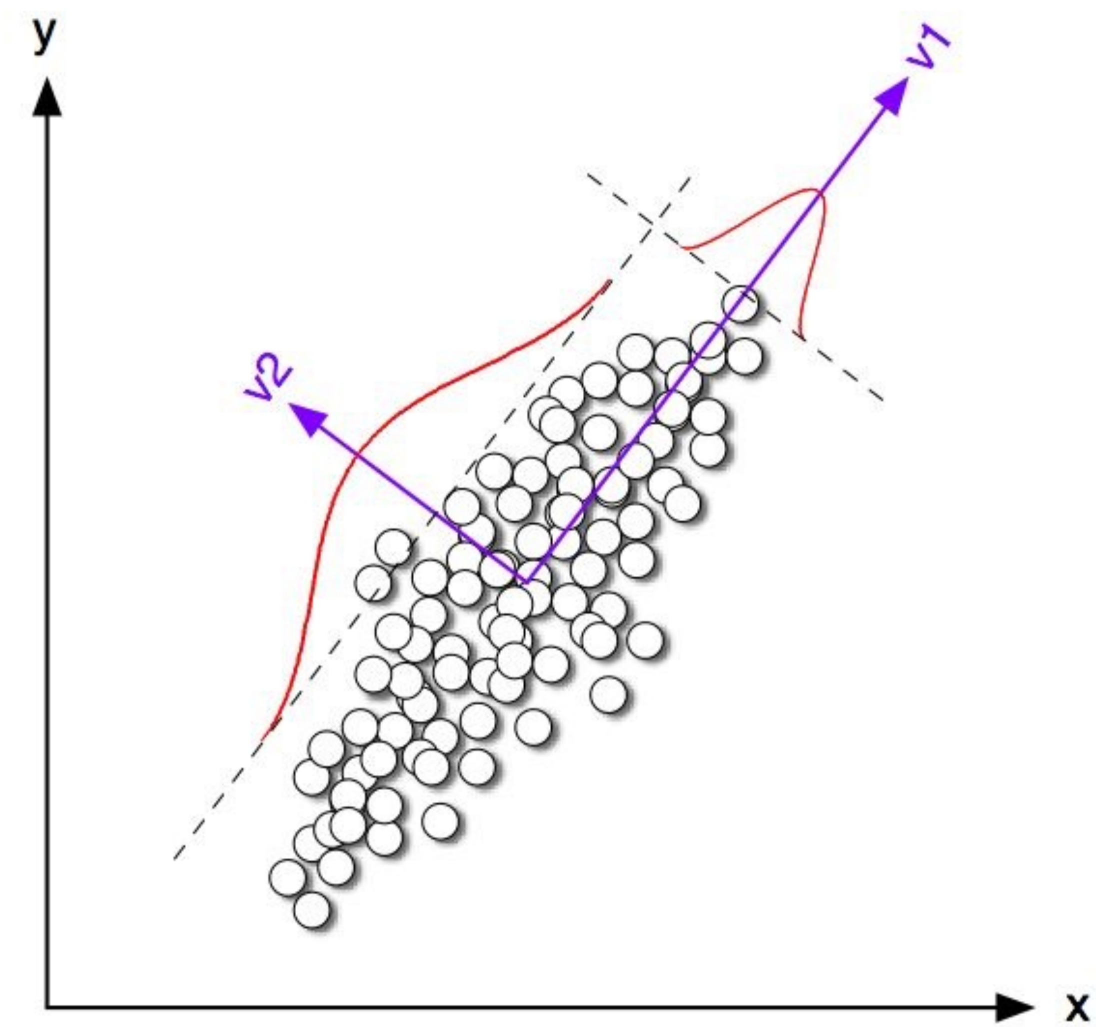
Augmenting dimensionality reduction

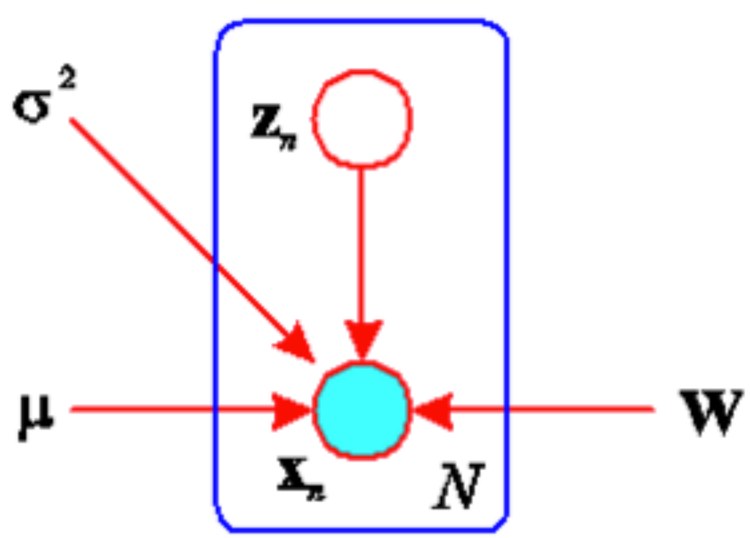
- When looking for a lower dimensional representation of your data, you typically enforce generic principles about the latent variables
  - independence, orthogonality, sparsity, etc.
  - these may correspond to stimulus features or experimental conditions, but not necessarily (especially given the structure of your model, e.g. linear mappings)
- Another approach is to put conditions explicitly in the model

Review of PCA

- goal: explain as much variance with as few variables (components) as possible

- parameter estimation: eigenvectors of data covariance matrix
  - or an iterative Expectation Maximisation parameter estimation algorithm





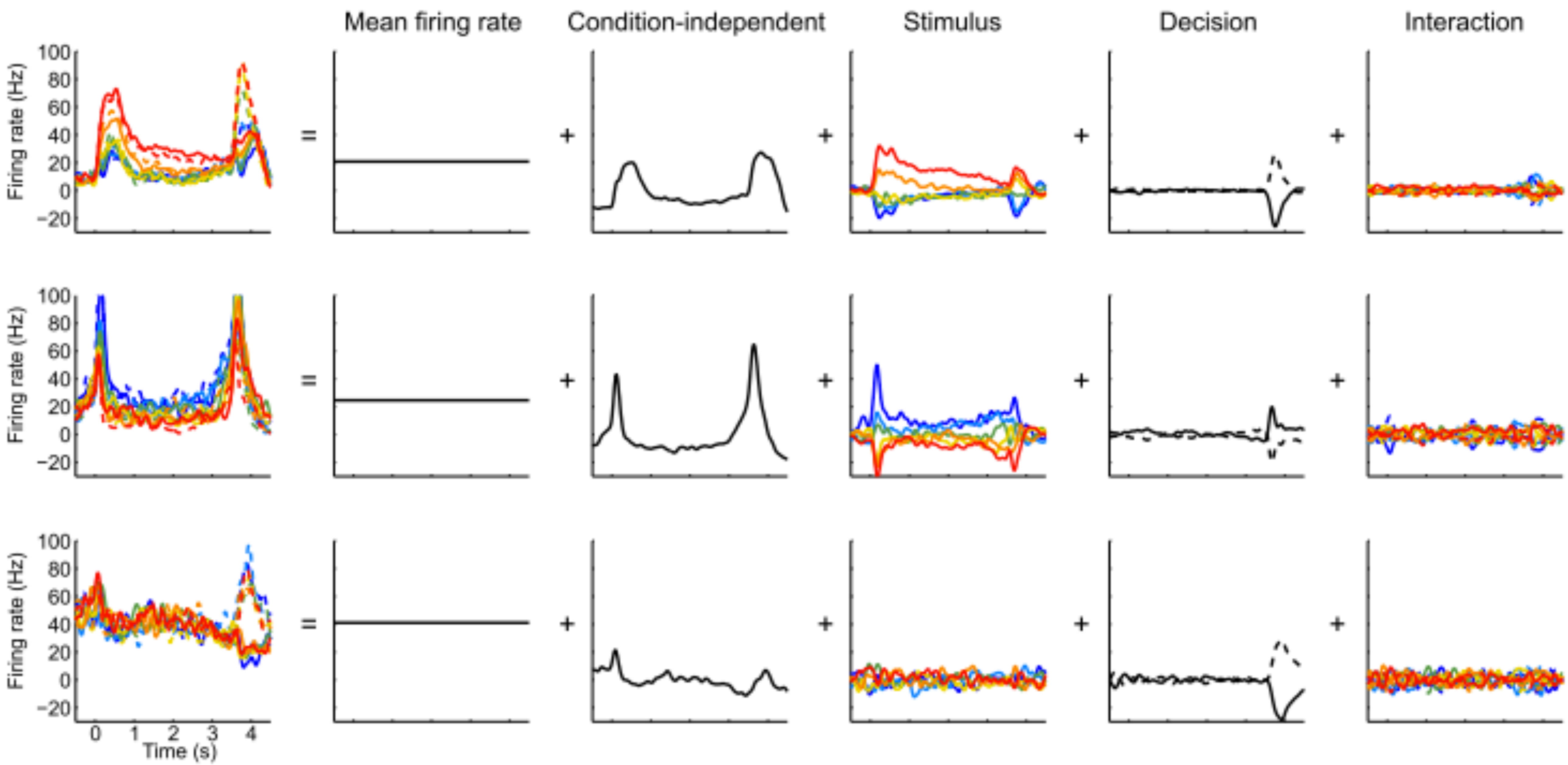
$$p(z) = \mathcal{N}(z; 0, \mathrm{I})$$

$$p(x \mid z) = \mathcal{N}(x; \mathrm{W}z + \mu, \sigma^2 \mathrm{I})$$

Data

- four different datasets from the labs of Romo, Constantinidis and Mainen
- monkey PFC, rat OFC
    - decision tasks, in each trial there is a stimulus (tactile, visual or olfactory) and the animal makes a binary decision
- they always use the peristimulus time histogram as data
    - they average over trials (needed because cells are not recorded simultaneously, and have diff. number of trials too)
    - data dimension: #cells x (#stim · #dec · #timebin)

Demixing



Mean firing rate
Condition-independent
Stimulus
Decision
Interaction
Firing rate (Hz)
100
80
60
40
20
0
-20
=
+
0
1
2
3
4
Time (s)

Decomposition of variance

- If we decompose the data by marginalising over different parameters, the covariance matrix can be given as a sum of the decomposed covariances

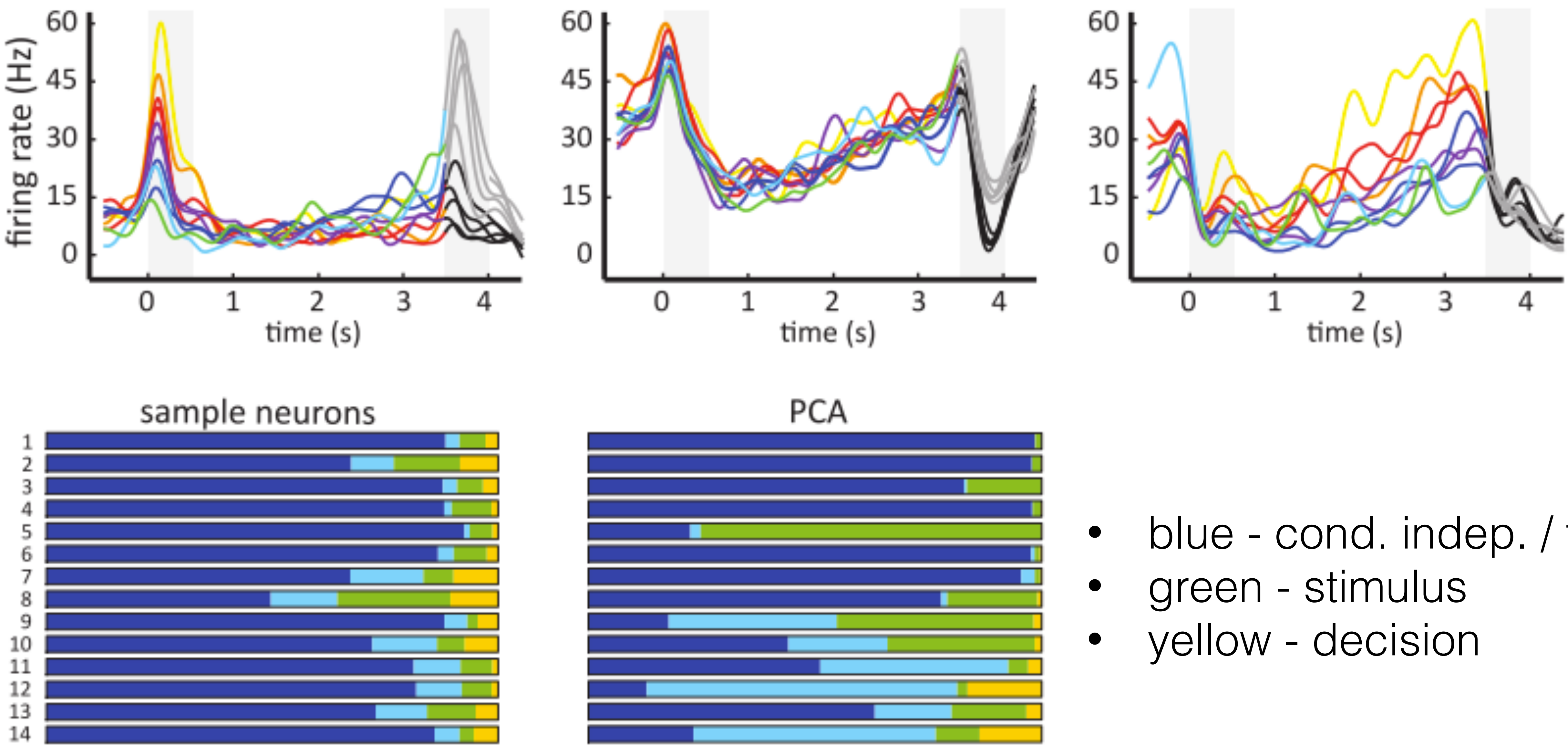
$$x(t,s) = \mu + z(t) + z(s) + z(t,s).$$

$$z(t) = \langle x(t,s) - \mu \rangle_s \,. \qquad z(s) = \langle x(t,s) - \mu \rangle_t \,. \qquad z(t,s) = x(t,s) - \mu - z(t) - z(s).$$

$$\mathrm{Var}(x(t,s)) = \mathrm{Var}(z(t)) + \mathrm{Var}(z(s)) + \mathrm{Var}(z(t,s)).$$

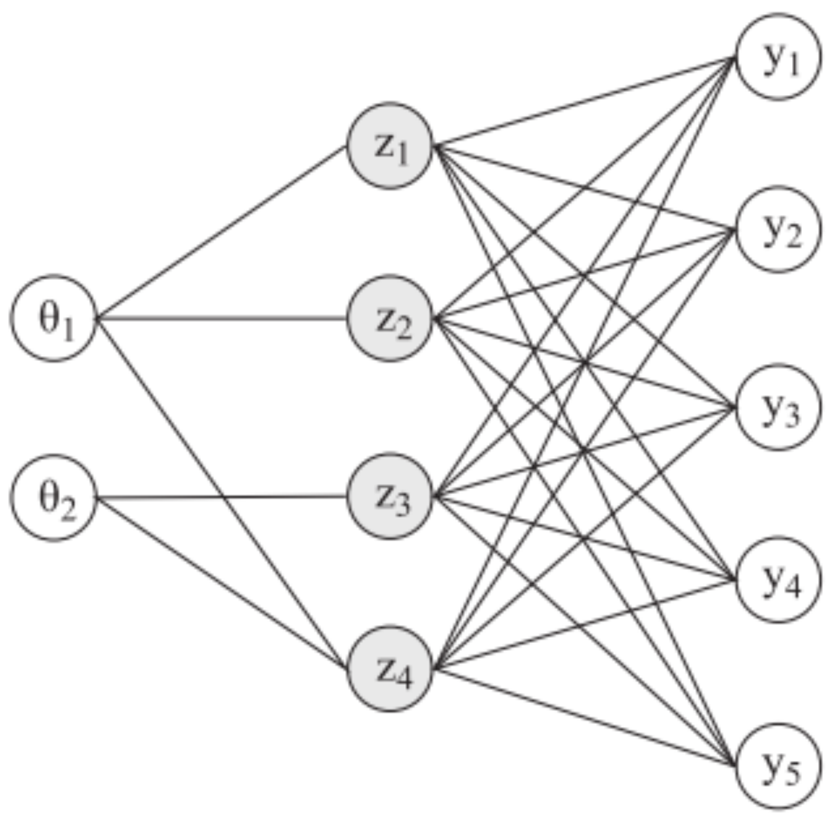
- Instead of regular principal components, we are looking for ones that describe variance only in the direction of the eigenvectors of one of the decomposed covariances

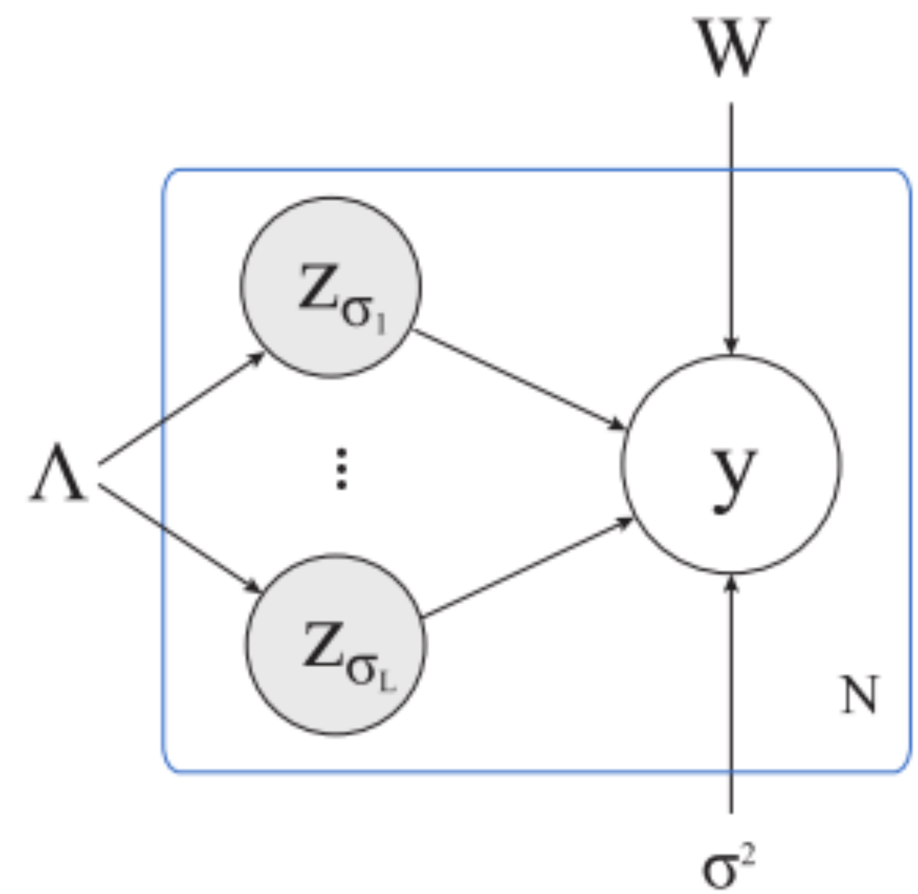
Selectivity of neurons and principal components



- blue - cond. indep. / time
- green - stimulus
- yellow - decision

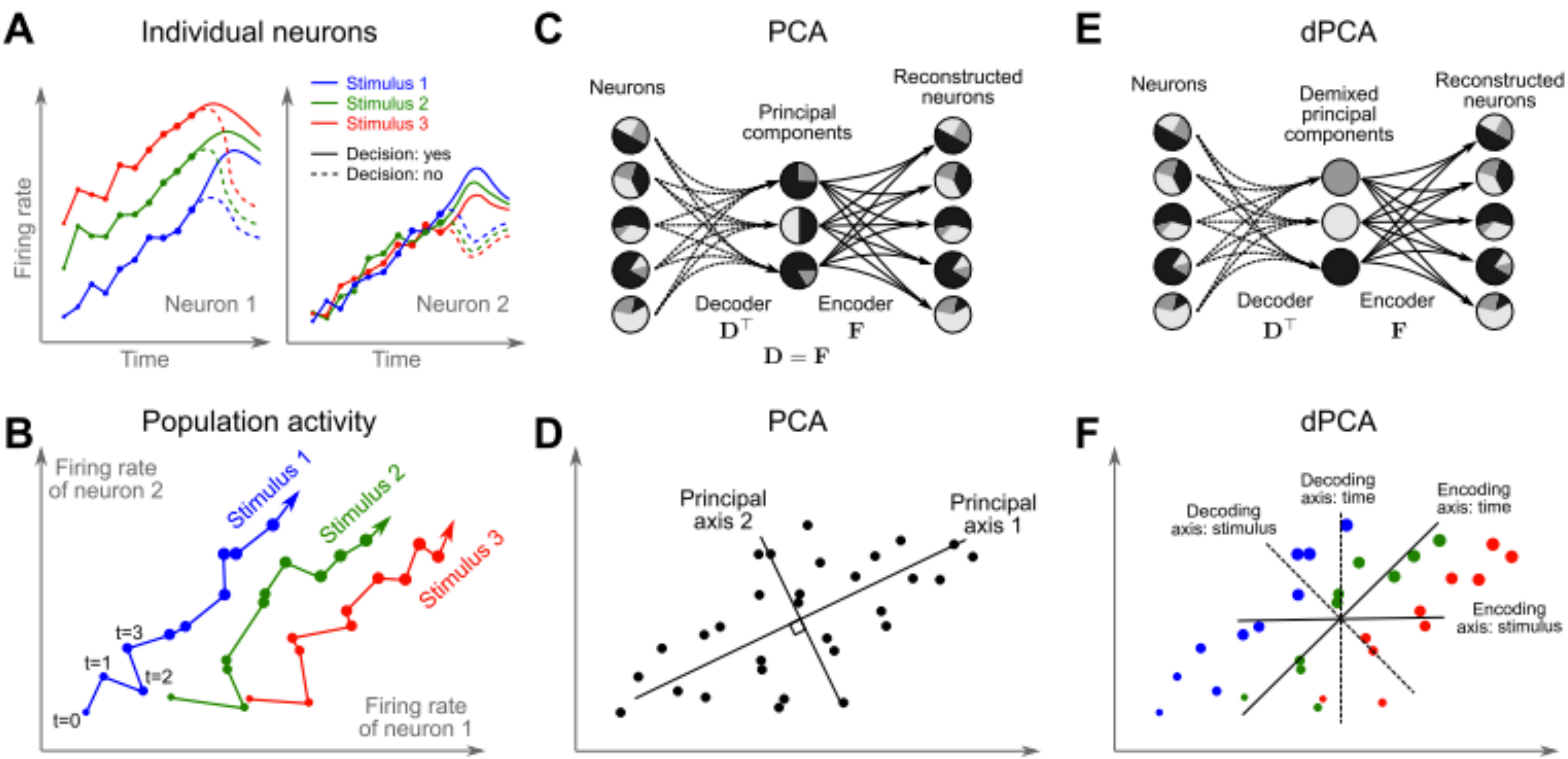
Generative model of dPCA

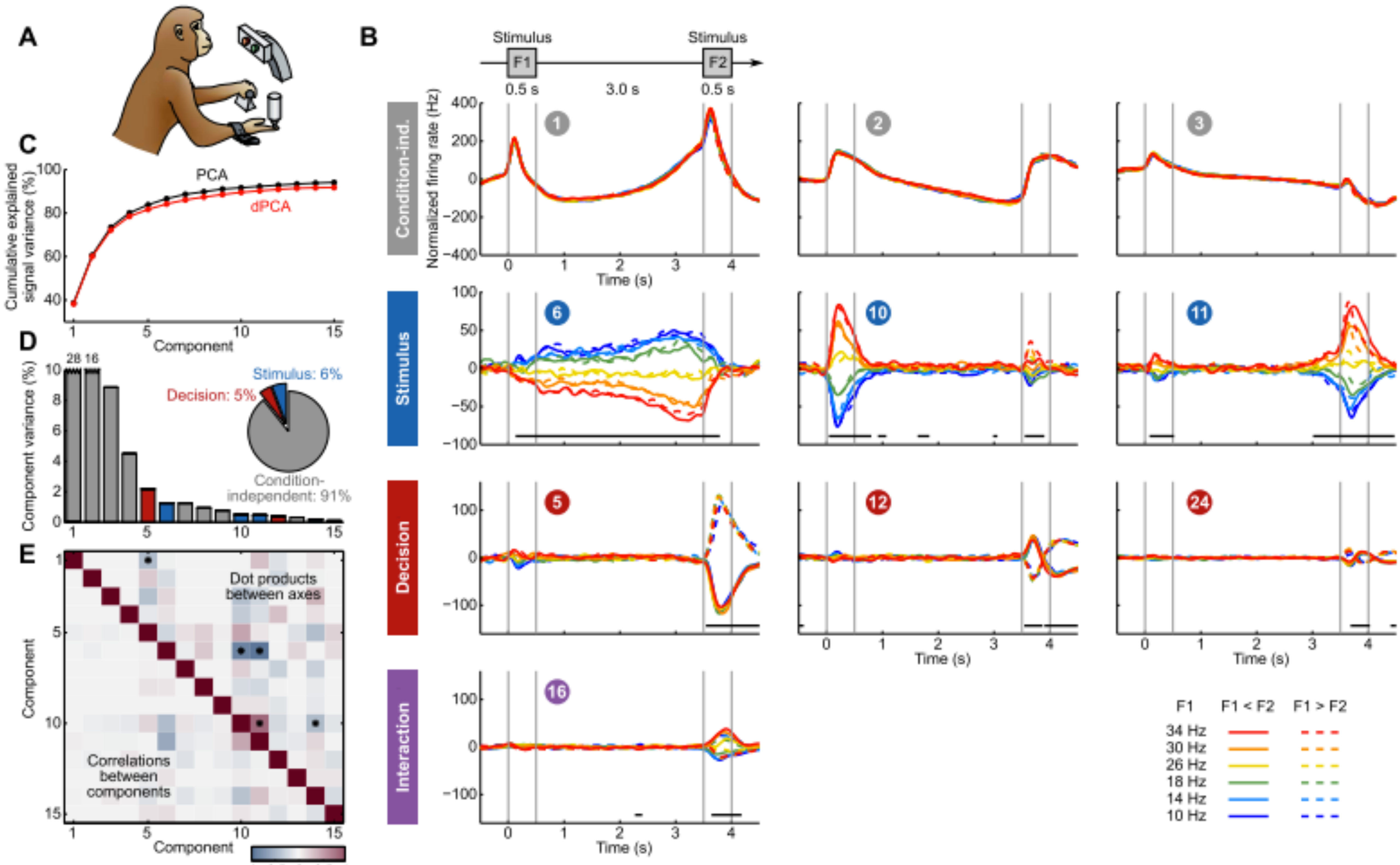




- An EM algorithm can be derived to find the components by adding constraints to the standard PCA inference

PCA vs. dPCA

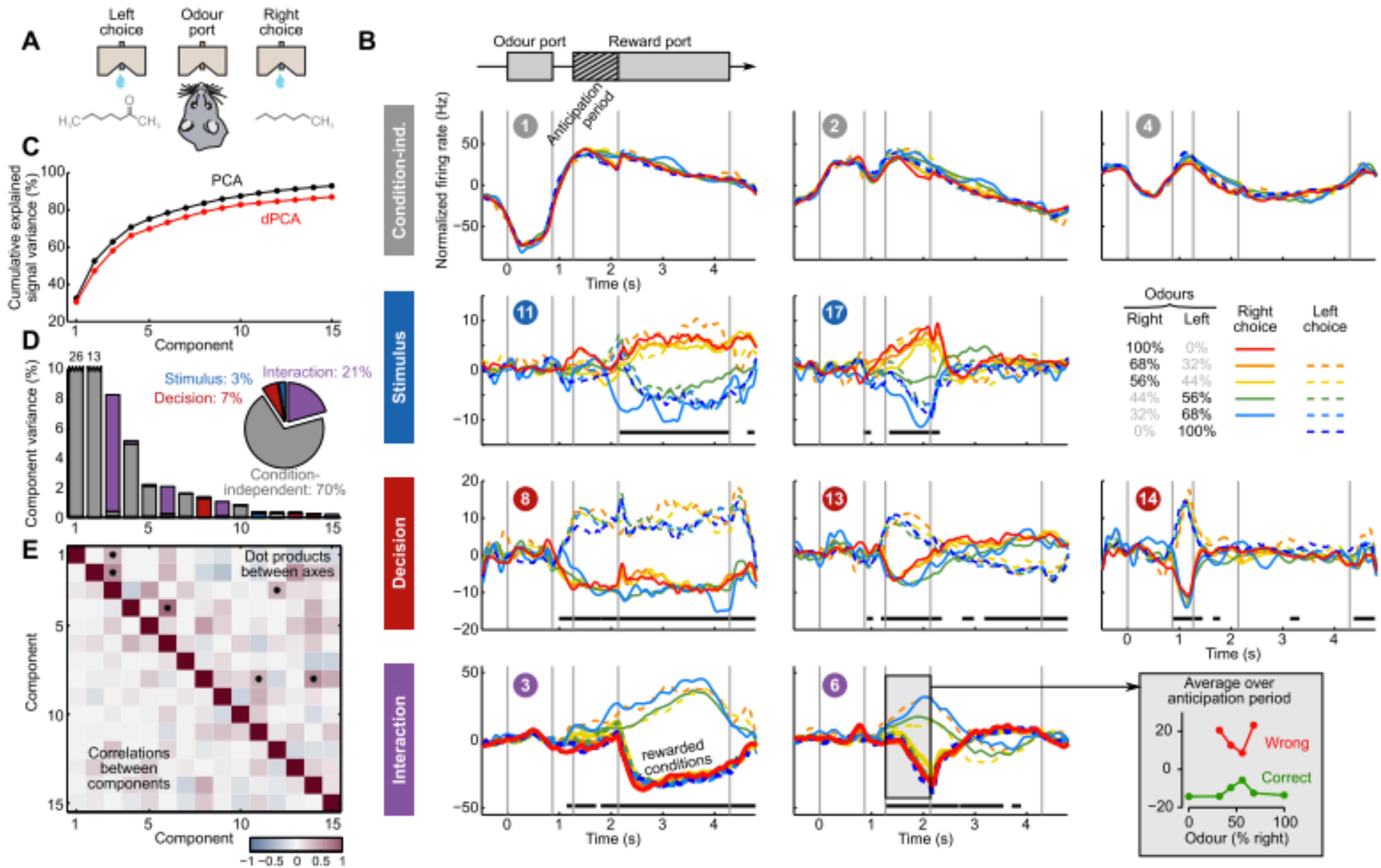

A
Individual neurons
Firing rate
Time
Neuron 1
Neuron 2
Stimulus 1
Stimulus 2
Stimulus 3
Decision: yes
Decision: no
B
Population activity
Firing rate of neuron 2
Firing rate of neuron 1
Stimulus 1
Stimulus 2
Stimulus 3
t=0
t=1
t=2
t=3
C
PCA
Neurons
Principal components
Reconstructed neurons
Decoder $\mathbf{D}^\top$
Encoder $\mathbf{F}$
$\mathbf{D} = \mathbf{F}$
D
PCA
Principal axis 2
Principal axis 1
E
dPCA
Neurons
Demixed principal components
Reconstructed neurons
Decoder $\mathbf{D}^\top$
Encoder $\mathbf{F}$
F
dPCA
Decoding axis: stimulus
Decoding axis: time
Encoding axis: time
Encoding axis: stimulus



A
B
C
D
E
Stimulus
Stimulus
F1
F2
0.5 s
3.0 s
0.5 s
Condition-ind.
Stimulus
Decision
Interaction
Normalized firing rate (Hz)
Time (s)
Cumulative explained signal variance (%)
PCA
dPCA
Component
Component variance (%)
Stimulus: 6%
Decision: 5%
Condition-independent: 91%
Dot products between axes
Correlations between components
F1
F1 < F2
F1 > F2
34 Hz
30 Hz
26 Hz
18 Hz
14 Hz
10 Hz

Insights

- Linear demixing of the population activity is possible
- Most of the variance is explained by components not related to any conditions
- Some components are the derivatives of each other -> similar activity patterns arise in the population with temporal shifts
- Using decision-related components as fixed linear decoders, one can decode the decision from the population activity with ~75% accuracy



A
Left choice
Odour port
Right choice
B
Odour port
Reward port
Anticipation period
Condition-ind.
Stimulus
Decision
Interaction
Normalized firing rate (Hz)
Time (s)
Odours
Right
Left
Right choice
Left choice
100%
68%
56%
44%
32%
0%
rewarded conditions
Average over anticipation period
Wrong
Correct
Odour (% right)
C
Cumulative explained signal variance (%)
PCA
dPCA
Component
D
Component variance (%)
Stimulus: 3%
Decision: 7%
Interaction: 21%
Condition-independent: 70%
E
Dot products between axes
Correlations between components
Component

Limitations

- Needs lots of neurons
- All parameter combinations need to be present in the data
- Applied only to PSTH here, in theory it might be applicable to trial-to-trial covariability, but not tested

Functional generative models vs. demixed dimensionality reduction

- Generative models of perception
  - neural activation / latent causes -> stimulus
  - emphasis on prediction of neural activity statistics
- Demixed dimension reduction
  - stimulus / conditions -> neural activation
  - emphasis on identification of variable mappings
- Hypotheses about neural representations are formalisable in both frameworks