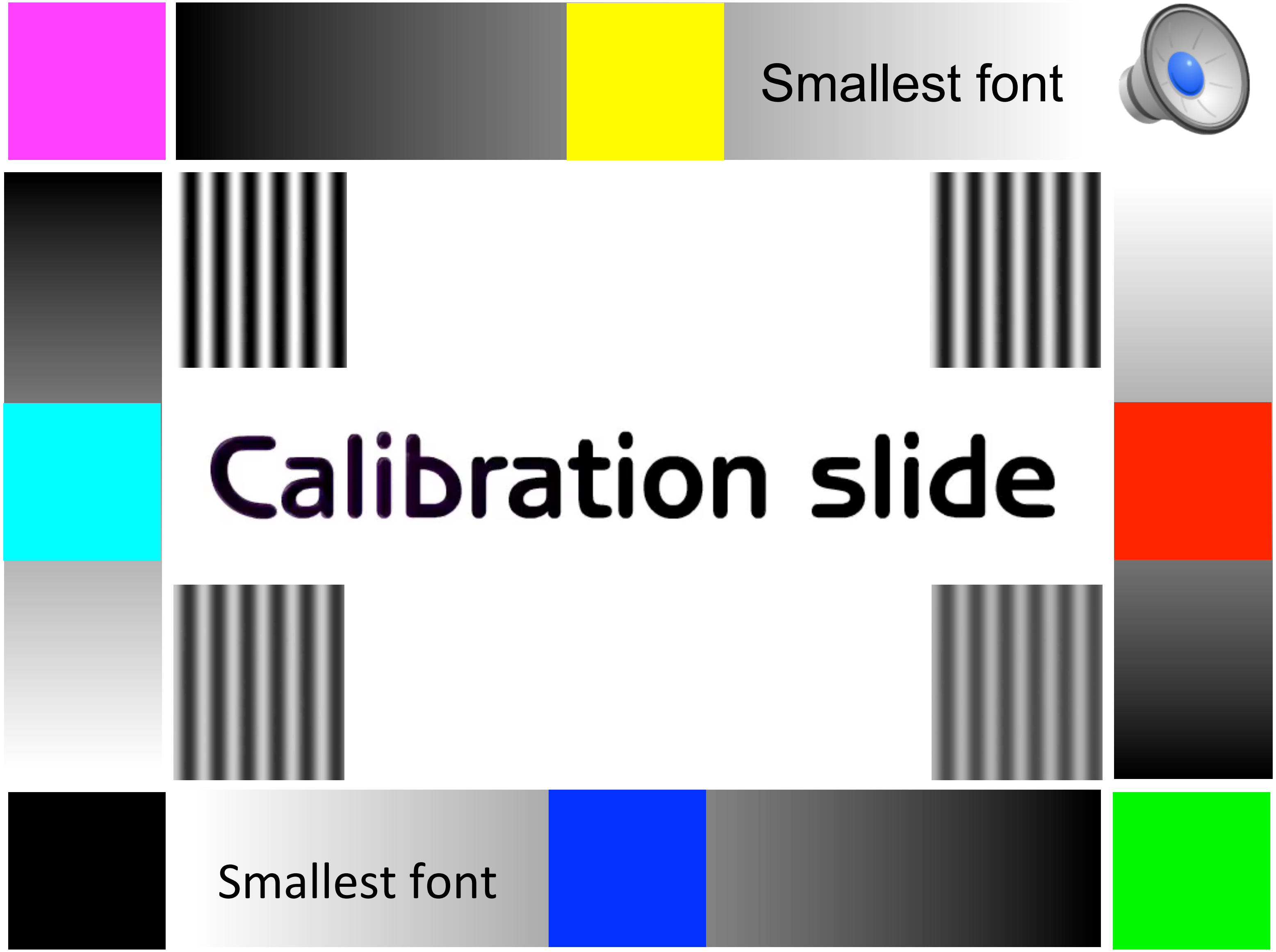


Smallest font

Calibration slide

Smallest font

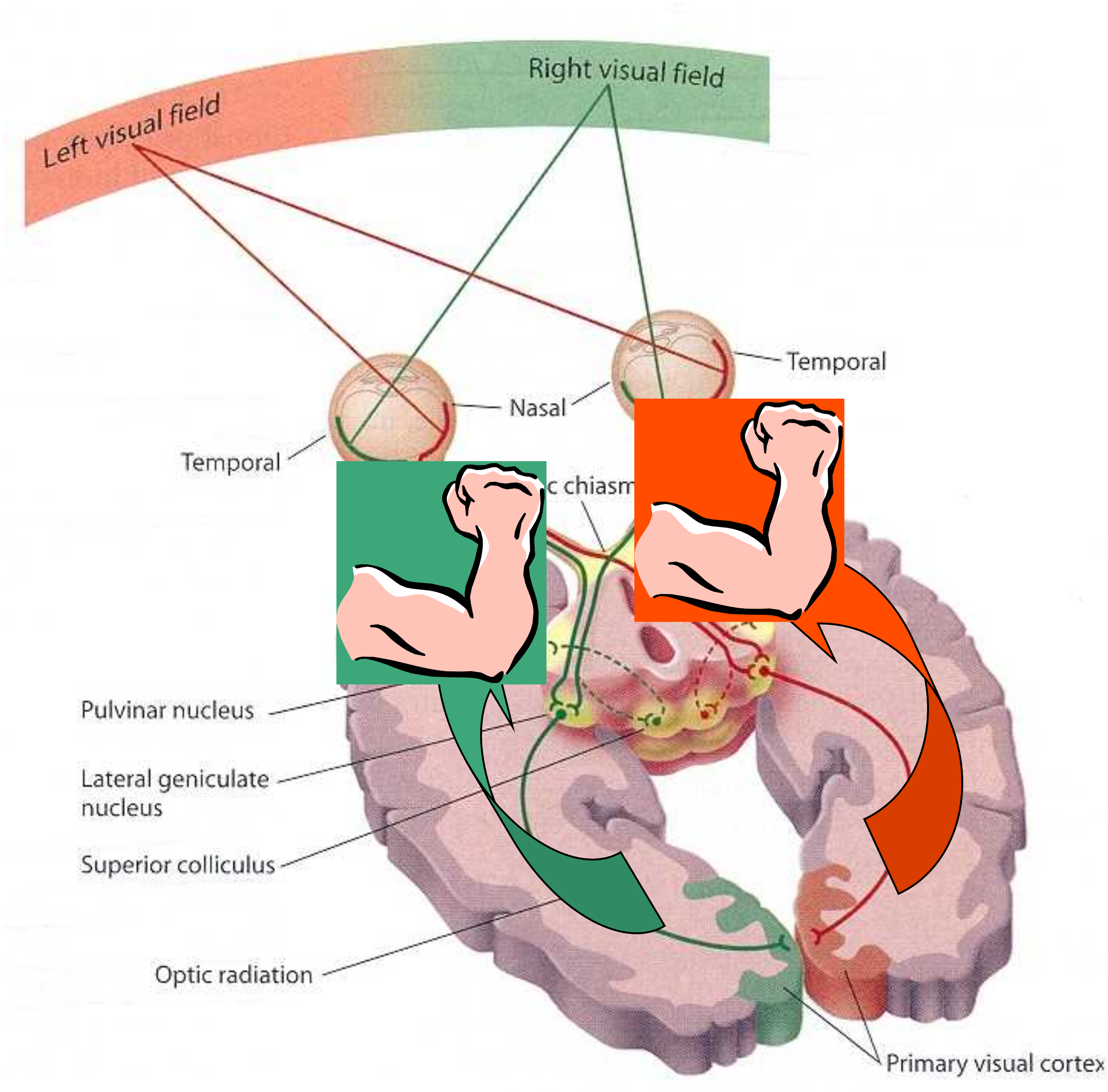
Principles for the analysis of (big) neural data

- Pascal Wallisch -

One of the biggest problems I faced as a grad student and post-doc

- How do I organize data analysis such that there is a clear logical flow from getting the data to producing stats and figures?
- In a way that is easy to maintain (there *will* be requests for additional analysis and revisions).

The visual processing cascade



1) The canonical data analysis cascade

1 **Loader** - This program puts data into Matlab format. Analogous to transducer: Retina or cochlea.

2 **Pruner/Integrity checker/Filter** - This makes sure that the data to be analyzed is usable. Analogous to filter function of LGN or MGN. Thalamus. Choking off irrelevant data.

3 **Categorizer**. Analogous to V1. Format the data properly.

4a/b/c **Calculator** Analogous to extrastriate cortex. Do specialized calculations on the same data. Specialized streams.

5a/b/c **Plotter** - Output. Analogous to motor cortex. Makes figures. Also saves files (memory).

6 **Wrapper** - invoking the previous 5 programs, in the correct order. That would be the brain itself. Heavily commented.

We recommend to model large scale data analysis programs to follow the steps involved in neural information processing

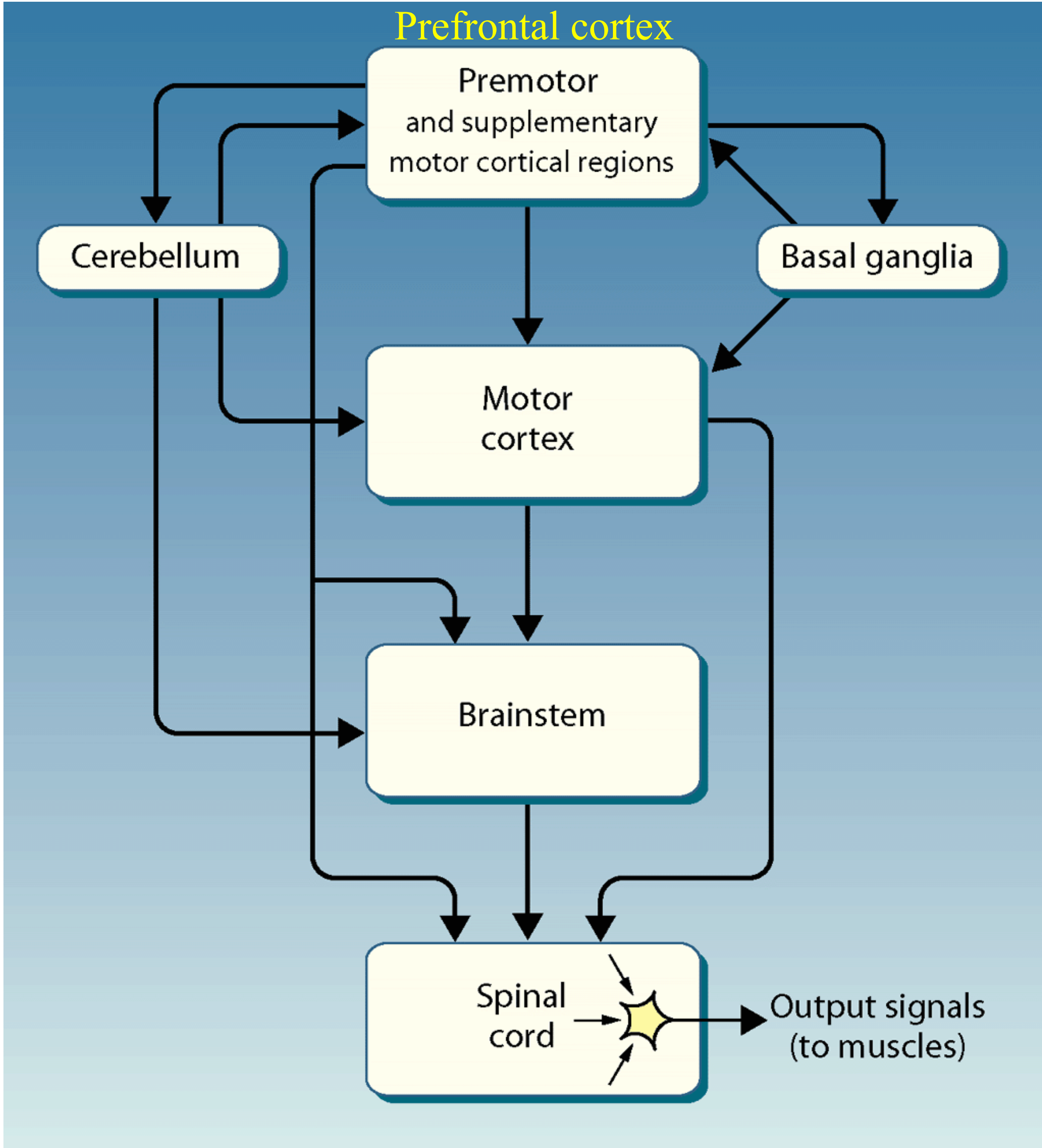
- There is another level to this.
- We *also* recommend to implement the hierarchical design principles of the primate motor system when engaged in scientific coding.
- After all, you are telling a computer what to do.

Neural basis of motor control hierarchy

Goal (movement selection & plan)

Tactics (spatio-temporal pattern of joint angles and muscle activations)

Execution


Prefrontal cortex
Premotor and supplementary motor cortical regions
Cerebellum
Basal ganglia
Motor cortex
Brainstem
Spinal cord
Output signals (to muscles)

2) **STRAGOTAI**:
The Coding implementation hierarchy

- **Stra**tegic **go**al: What are we trying to do (and why)?
- **T**actics/**A**lgorithm: How will we do it?
- **I**mplementation: What is the actual syntax?

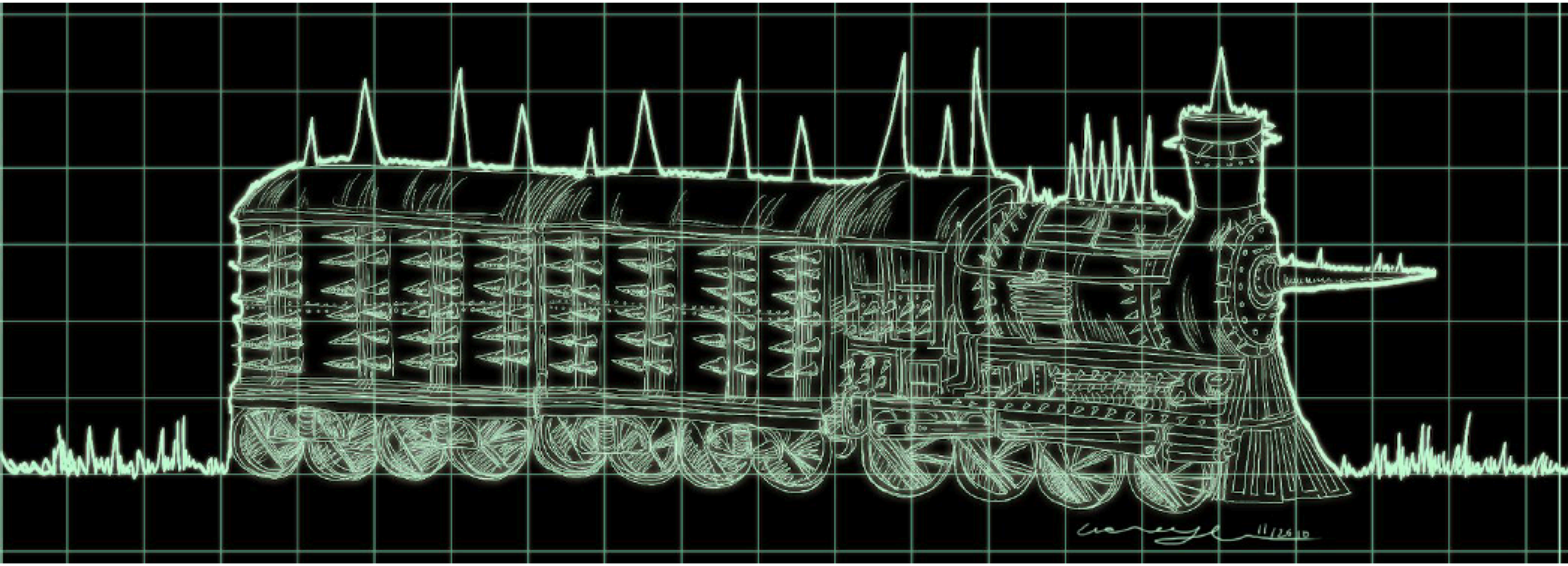
A paradigmatic case that will come up a lot

- Strategic goal: Dimension reduction because the data is too highly dimensional to make sense of, in the raw.
- Tactics: Factor Analysis, Principal Component Analysis (PCA), Independent Component Analysis (ICA), MAVE, …
- Implementation: Writing the actual (Matlab) code.
- Suggestion: Use this for comments hierarchy (big goal, pseudocode, explaining variables in line)
- This makes code understandable, maintainable and portable.

3) Exploratory data analysis

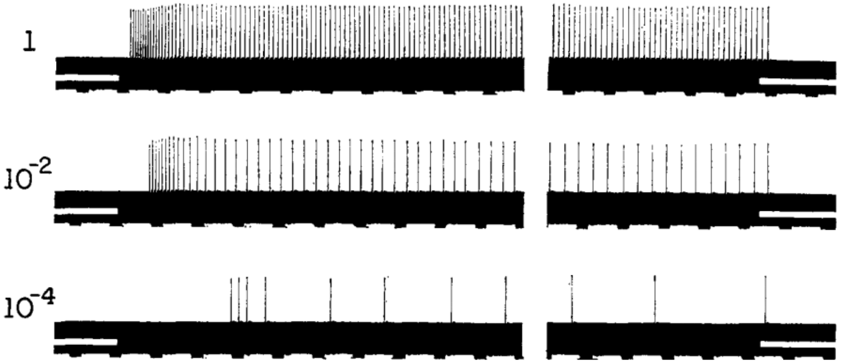
- Looking at *raw* data before processing it is absolutely critical, indeed necessary.
- Any data processing makes assumptions and transforms – information – potentially in a distorting fashion unknown to the analyst.
- Just like with food, the end result of highly processed data can be unbecoming.

Characterizing spike trains

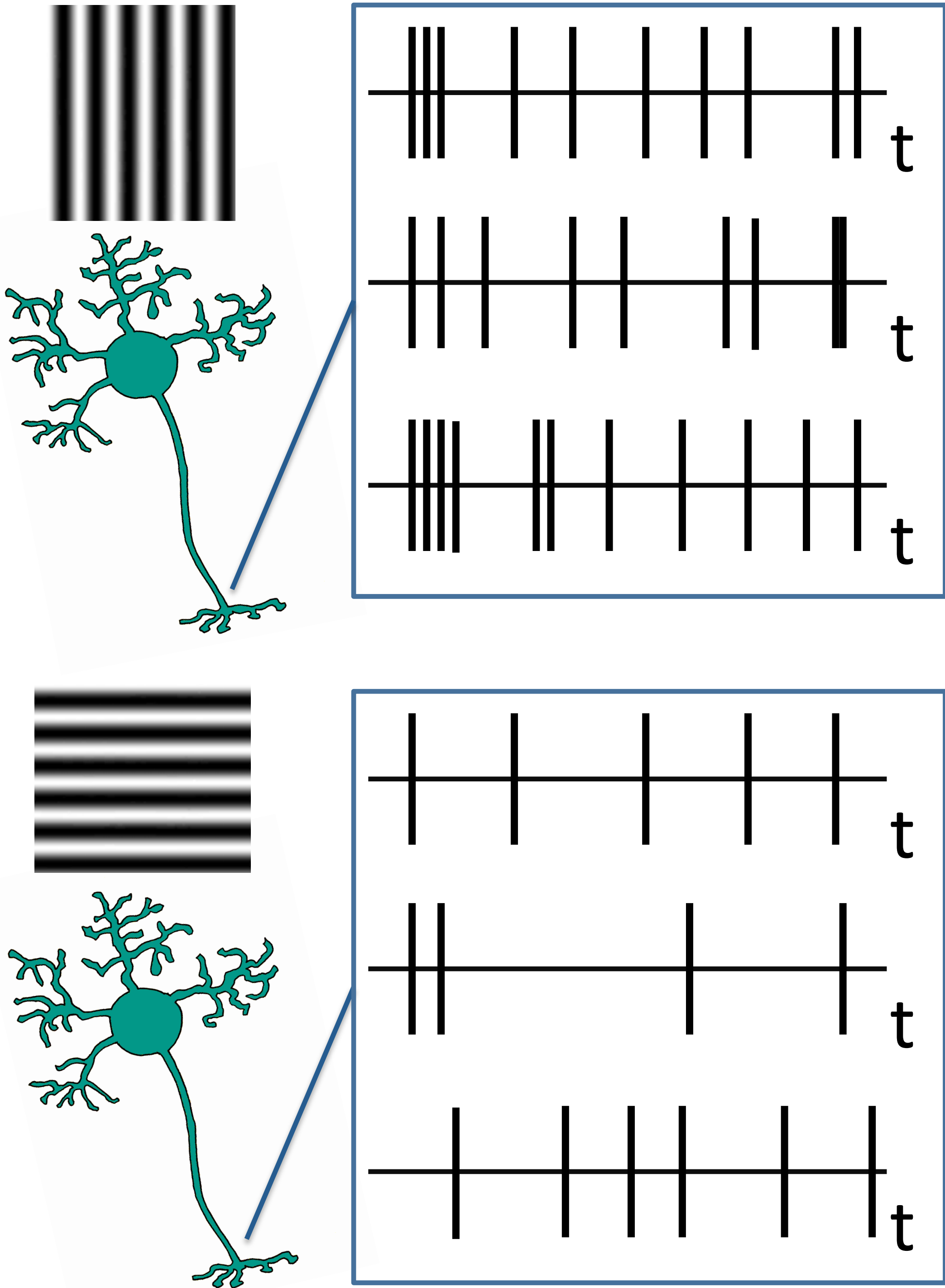


- Raster plots
- Peri-stimulus time histograms (PSTHs)
- Tuning curves

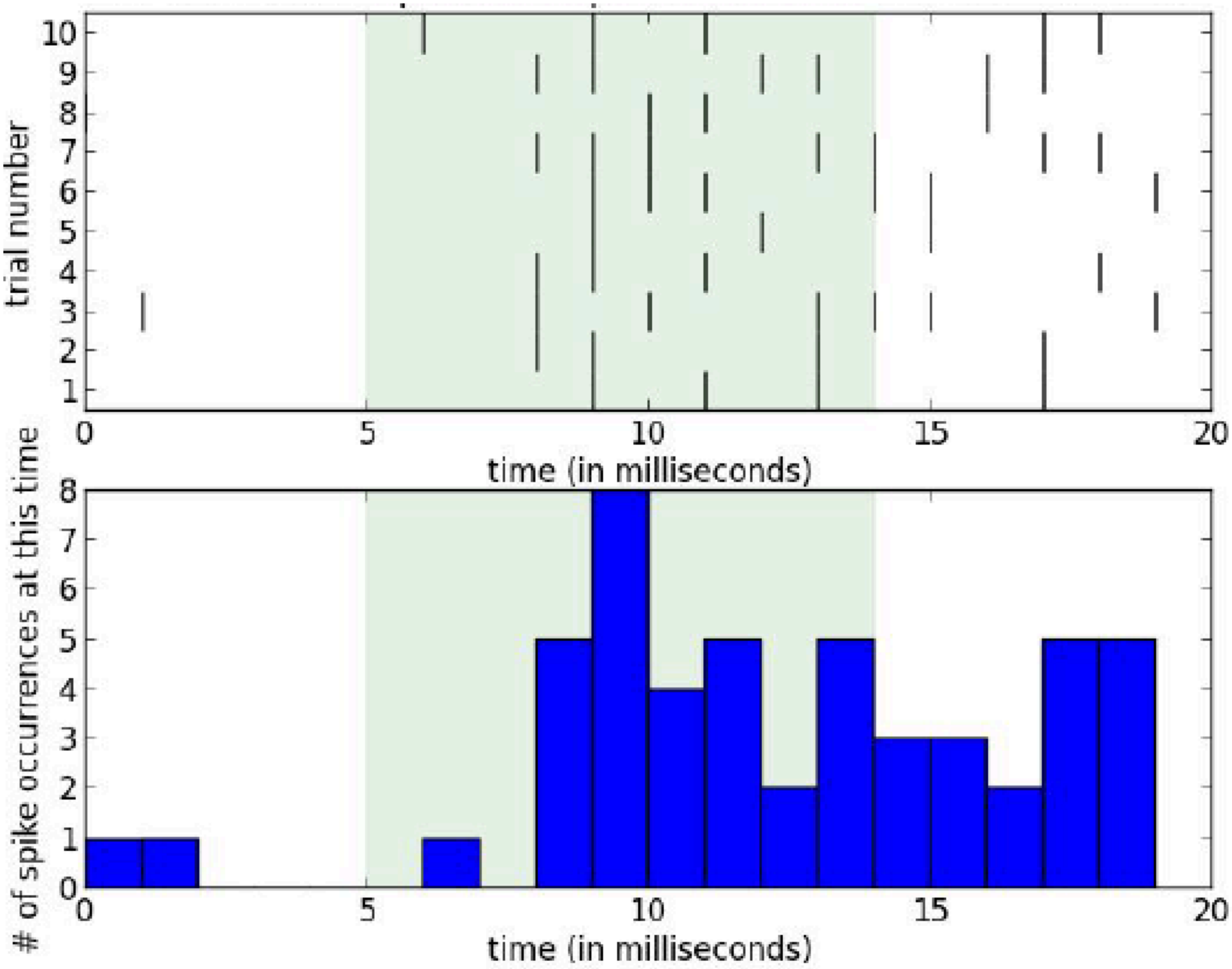
Raster plots



HK Hartline
(Nobel 1967)



From raster plots to the PSTH: The peri-stimulus time histogram



Some final “codmandments” 😬

- Each logical paragraph of code maximally fit on one screen in the editor (if it takes up more than that, it will be hard to maintain; instead, use functions)
- Code from the “inside out”, make sure one trial of one condition works, then all trials from one condition, then all conditions, then all participants, etc.
- Don’t trust, verify: Test that the code does what you think it does, never assume it does
- Separate storage structures from computing structures
- Control (“seed”) the random number generator (!)
- Use “camel case” when naming variables
- …