COGS138: Neural Data Science

Lecture 17

C. Alex Simpkins, PhD
UCSD Dept. of Cognitive Science, Spring 2023
RDPRobotics, LLC
http://casimpkinsjr.radiantdolphinpress.com/pages/cogs138_sp23

rdprobotics@gmail.com | csimpkinsjr@ucsd.edu

Plan for today

- Announcements
- In class paper
- Single trial analysis introduction, examples and readings

Announcements

- ***<u>Deadlines upcoming this week:</u>***
- **Tuesday**:
- **Wednesday:**
- **Saturday**:
    - Reading Quiz 4 11:59pm
    - In-class paper completion and submission
    - Lecture quiz
- **Friday**:
    - Data checkpoint 11:59pm

Announcements

- We are working on getting the canvas grades updated with the weights, so this week you will be able to check in on that
- Project meetings went well, we are reviewing the proposals and will provide direct feedback in the github repos as ‘issues’
- About final presentations
    - Depending on class size we usually with such classes either have groups create slides and record a video presentation or
    - In class presentations - more educational for everyone seeing all the projects, techniques, issues and conclusions, and we have a small enough class but want your input as part of it

Announcements

- **github** repos
  - created,
  - invites sent,
  - please accept (time limited, most have)
  - login and be sure you can and files are there, rename
- if you don't have an invite, there's an issue with your group record in the main list - please contact us asap
  - **Procedure** : Contact Siddhant, cc me, if no response in a day, reach out to me again, I'll help

Project schedule

Task due	Date due	Description
Previous project review	5/23/2023 at 11:59pm (Tuesday)	Select 2 of the 3 available, review as individuals and then come together as a group to submit your responses to the questions after a discussion. This will orient you to the class project
Project proposal	5/26/2023 at 11:59pm (Friday wk8)	Generate your question, hypothesis, initial data sets you'll be working with, etc., describe your plan, schedule, who is doing what, potential issues, suggested analysis and how it will answer your question
Data checkpoint	6/2/2023 at 11:59pm (Friday wk9)	Builds on the proposal by taking the feedback from PP above and actually getting, loading, describing your data,
EDA checkpoint	6/9/2023 at 11:59pm (Friday wk10)	Builds on the previous checkpoint, essentially most of your analysis should be done by this point
Final report	6/15/2023 at 11:59pm (Thursday Fin wk)	Due Thursday of finals week so we can grade before the Tuesday deadline, otherwise your grade may be delayed
Group evaluations	6/15/2023 at 11:59pm (Thursday Fin wk)	You will evaluate each other based on participation and performance, this will contribute to your overall final project grade 5%)

Remaining assignments schedule

- A4 wk9-10, A5 extra credit
- R4 wk9
- LQquiz wk 9, 10
- Paper completion this week, mostly in class or via appointment

Last time…

Parameterizing heterogeneous datasets

- Definition, review
  - What do we mean by **parameterization**?
  - Reminder of what data is and stepping back to the big picture - ***representation***
  - What are **heterogeneous** datasets?
  - What are the **challenges** and solutions?
- **Tools and practice** in neural data science
  - https://nwb-overview.readthedocs.io/en/latest/tools/tools_home.html
- Examples

Parameterization vs. Hyperparameterization

- **Parameterization** - the set of parameters that define the model unknowns to be fit, typically from data
    - For example, for $y = ax + b$, what are the parameters?
    - ANN - network weights
    - Calculated/learned from data
- **Hyperparameterization** - the set of parameters for machine learning in particular that define and control the learning process and are external to the model
    - Bisection algorithm for optimization - bisection parameter
    - ANN - parameters of the learning algorithm itself
    - Heuristic, can be set by practitioner, tunable for a given problem

Parameterization vs. Hyperparameterization

- **Parameters**
  - Calculated/learned from data (“the fit”)
  - Internal to model
  - Chosen as part of model structure either manually or algorithmically
- **Hyperparameters**
  - Heuristic, can be set by practitioner, tunable for a given problem
  - External from model
  - Optimal parameters are not known, and are different for different problems
  - Other ex.: ANN learning rate, gradient descent step size, the *k* in *k*-nearest neighbors

Stepping back: What is data?

Stepping back: What is data?

- **Data** can be of many forms
- **Data** - any representation of information that has been recorded in a fixed or dynamic state [Simpkins, 2023]
- ***"(1) Factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation" [Webster's, 2023]***
- ***(2) "Information in digital form that can be transmitted or processed" [Webster's, 2023]***

What do we mean by ‘representation’

What do we mean by ‘representation’

- Some sort of arbitrary symbolic link between the reality and the symbols we use to model reality, such as words, numbers, pictures, graphs, sounds, videos, smells, textures, vibrations, gestures
- Further information: See Simpkins COGS100 lecture 10 on representation, and read Norman ch3: “The power of representation”

Representation defined

- Cognitive age, Norman argues started when we started using sounds, gestures and symbols to refer to objects, things and concepts - when we started generating data!
- **<u>Representation</u>** : The sound, gesture, symbol is not the thing itself, it stands for, refers to it
- On representation not the reality

Powers of cognition come from abstraction and representation

- Ability to represent perceptions, experiences, thoughts in some medium other than what they occurred in

- Abstracted away from irrelevant details

- “The essence of intelligence” as he states - if representation is just right, new experiences, insights, creations emerge

- **We can make symbols then use them to do our reasoning**

Representing the dimensions requires different types of data entirely

- Ultimately in neural data science we are reasoning about the brain and behavior, how it’s all interconnected and the dynamics of it
- Data makes it possible to reach beyond our immediate cognitive limitations to operate on information
  - We cannot see a neuron firing when we look at each other, we measure, but then must do something with that data, related it and connect it meaningfully to other things
  - As we have been reasoning, we need massive amounts of connections to understand the patterns of it all
  - Recording it all the same way often is impossible
    - EEG vs. Behavior, text, other dimensions

Data Structures Review

Structured data

- Can be stored in database SQL
- Tables with rows and columns
- Requires a relational key
- 5-10% of all data

Semi-structured data

- Doesn't reside in a relational database
- Has organizational properties (easier to analyze)
- CSV, XML, JSON

Unstructured

- Non-tabular data
- 80% of the world's data
- Images, text, audio, videos

(Semi-)Structured Data

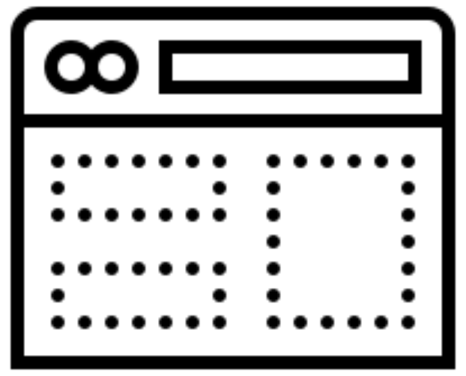
Data that is stored in such a way that it is easy to search and work with. These data are stored in a particular format that adheres to organization principles imposed by the file format. These are the data structures data scientists work with most often.

Unstructured Data

Some datasets record information about the state of the world, but in a more heterogeneous way. Perhaps it is a large text corpus with images and links like Wikipedia, or the complicated mix of notes and test results appearing in personal medical records.

Unstructured Data Types

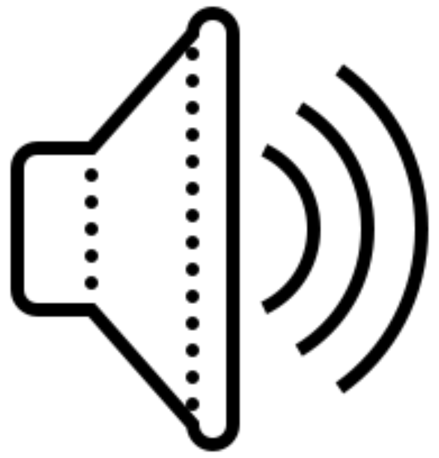
Text files and documents



Websites and applications

Sensor data

Image files



Audio files

Video files

Email data

Social media data

Source: TechTarget

What are heterogeneous datasets?

- Given that ***data can represent anything that can be represented***, we can have many forms of sampling and recording systems

- What have we covered thus far for data types and forms?

- Others?

- MOCAP
- EEG/MEG
- fMRI
- Eye tracking
- Text
- Single unit recording

Why integrate them?

- More information can draw links that may not be clear otherwise
- Limited data source sets may not contain the necessary data for the question we want to ask
  - **Sparsity** - improved results with ***sparse*** datasets
  - **Modality** - one set might have patterns, but lack the content explaining patterns, the meaning underlying
  - **Reliability** - one dataset showing statistical significance vs. many confirming from various perspectives
  - https://www.sciencedirect.com/science/article/pii/S1053811914003838
  - https://www.sciencedirect.com/science/article/pii/S1053811919300497

Why is it a challenge to integrate them?

- Sampling rate mismatch
- Time/frequency/spatial domains - what is the best form of representation?
    - https://www.sciencedirect.com/science/article/pii/S1053811919300497
- Sample rate variability (why does this matter?)
- Sample time mismatch
- Format, software
- Missing data, data mixture/non-tabular etc
- Memory usage
- (Not an exhaustive list)

Integration strategies - *Sample rate mismatch*

- Resampling
  - Sub-sampling - (“down-sampling”) - every Mth sample, LowPass first (aliasing)
  - Super-sampling - (“up-sampling”) - padding with 0’s, then LowPass to interpolate
- Interpolation/extrapolation (what are the differences?)
  - <u>Linear (LERP, BERP, TERP, SLERP)</u>
  - Piecewise continuous
    - <u>Splines, Bezier</u>
  - Polynomial
    - <u>Lagrange</u>, etc

Integration strategies - Time/frequency/spatial domains

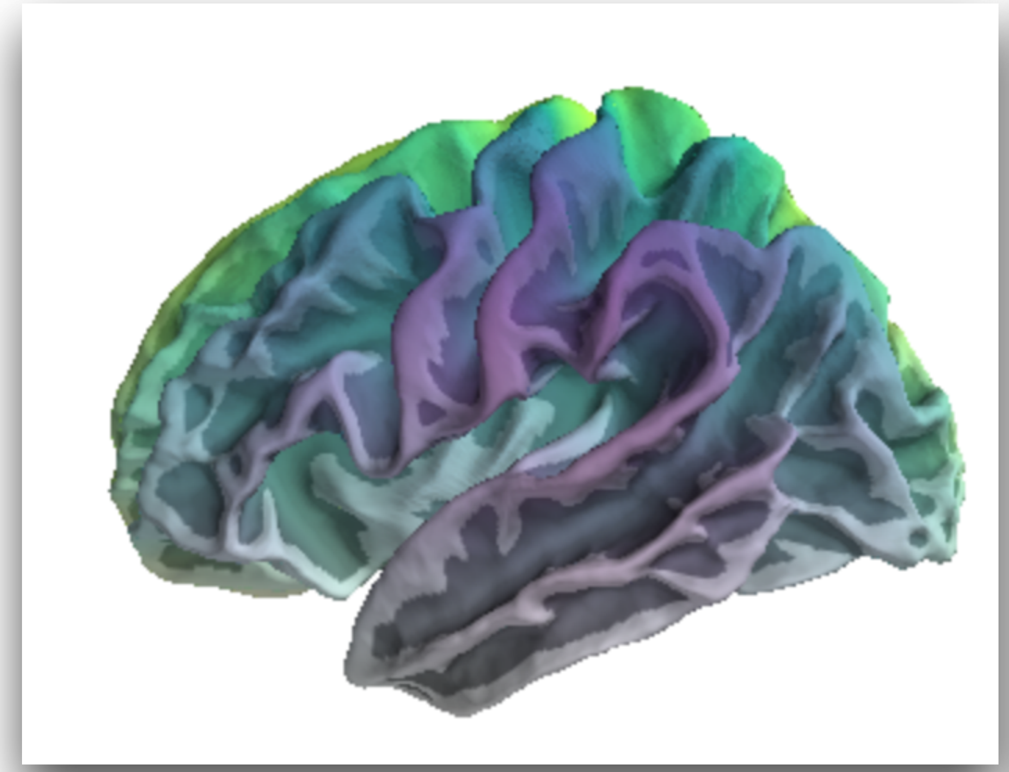
- We have data types such as structural scans of neural structure, EEG, MEG, fMRI, etc.

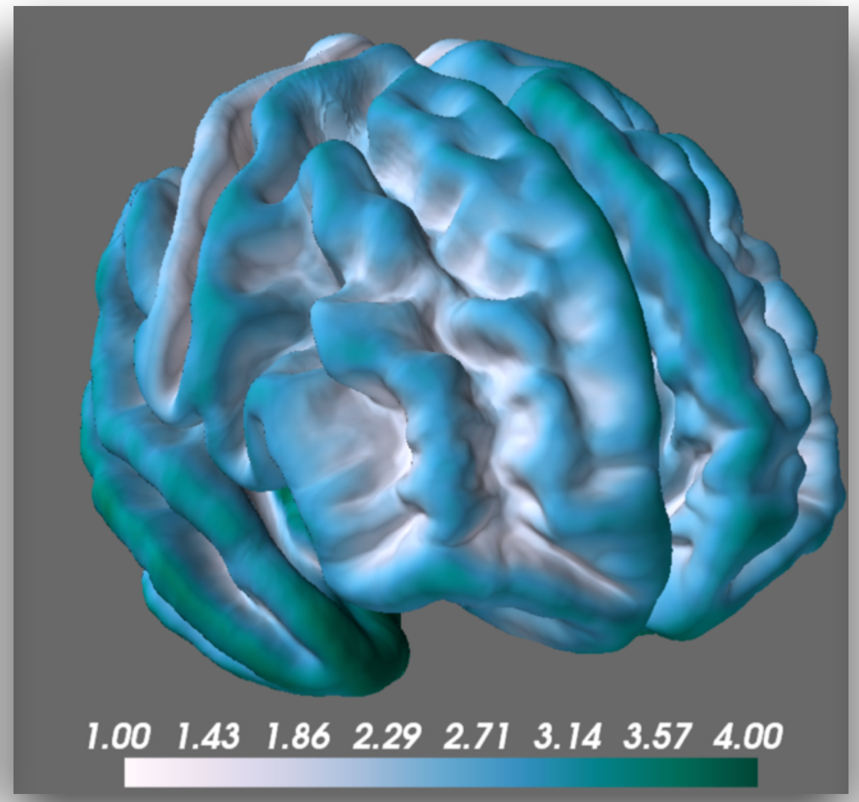
- How can these be synchronized spatially and temporally?

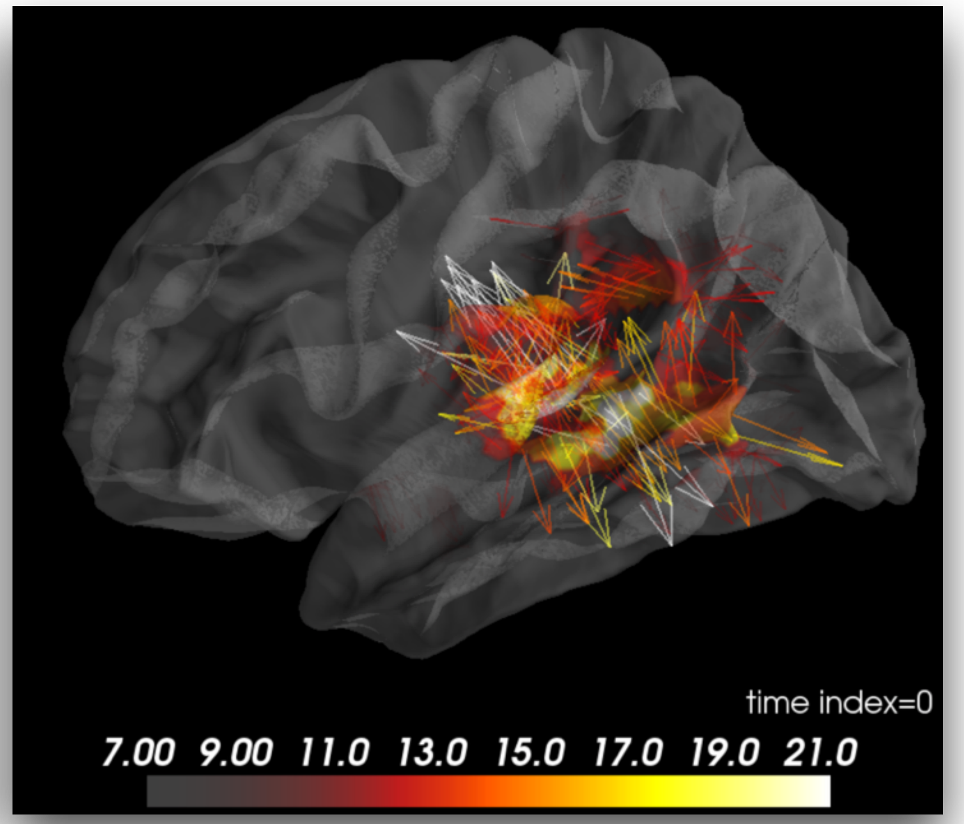
- What is an issue with spatial correlations (See A4!)?

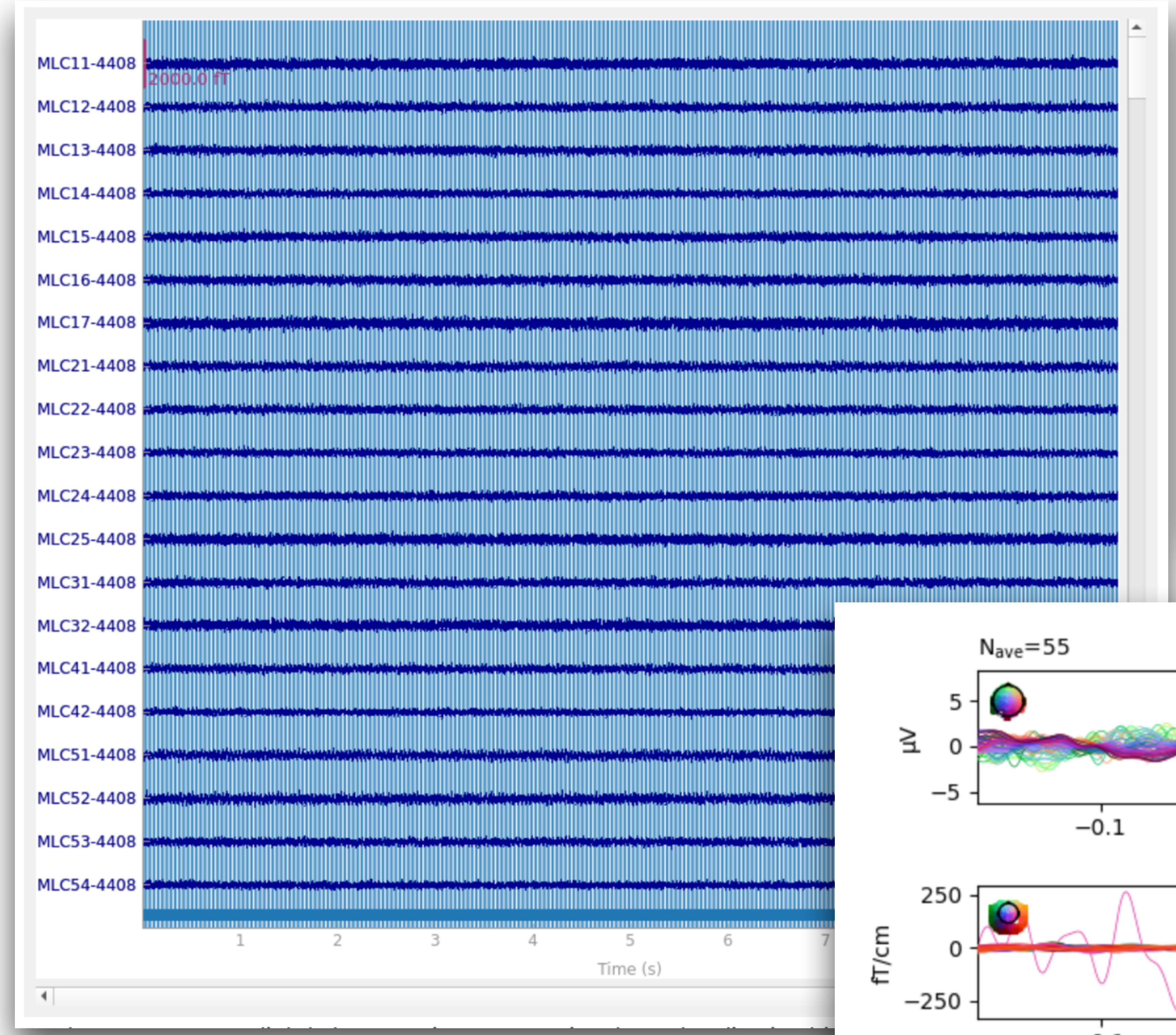
- Mapping - coordinate, typically affine transformation

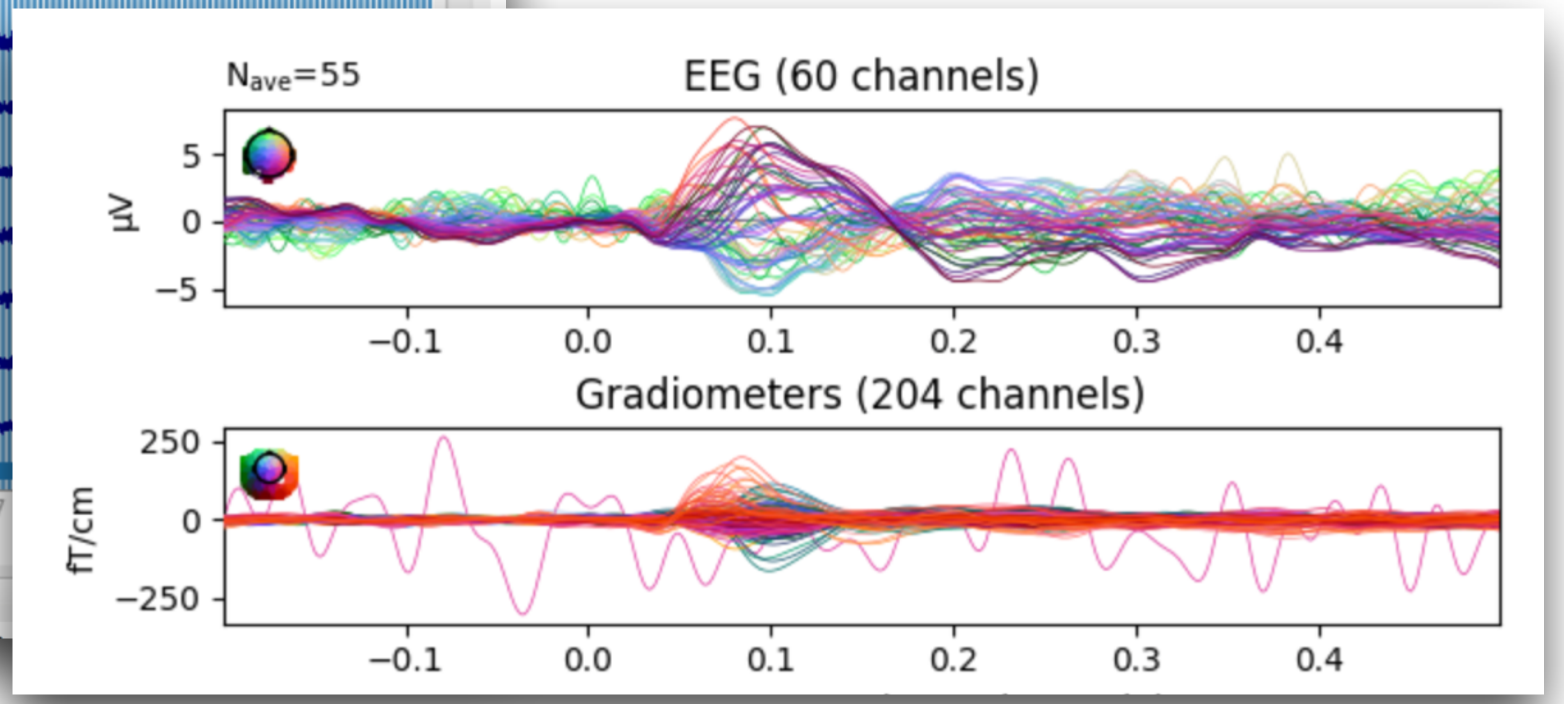
- Inverse computations - knowing locations of sensors relative to brain, can infer activation areas (localize)











Affine vs. Linear

- Can somebody explain the difference between ***linear*** and ***affine*** transformations?

- Requirements of linearity?

More on linearity vs. nonlinearity

- Power
  - **A linear system is a system whose dependent variables are related to its independent variables by a power of one**
- Linear systems have these particular properties (and they are very favorable)
  - **Additive** $\boxed{T[x_1(n) + x_2(n)] = T[x_1(n)] + T[x_2(n)]}$
  - **Homogeneous** $\boxed{T[cx(n)] = cT[x(n)]}$
  - (https://mathworld.wolfram.com/LinearSpace.html, https://mathworld.wolfram.com/LinearTransformation.html)

Affine transformation

- Any transformation that preserves collinearity (i.e. points on a line remain on a line after the transformation) and ratio of distances (midpoint of a line before and after transformation remains the same
- y=mx+b is? ***Affine***
- y=mx is? ***Linear***
- or more generally (see https://mathworld.wolfram.com/AffineTransformation.html, https://mathworld.wolfram.com/AffineSpace.html, https://medium.com/mlait/affine-transformation-image-processing-in-tensorflow-part-1-df96256928a)

Affine transformations in neural imaging

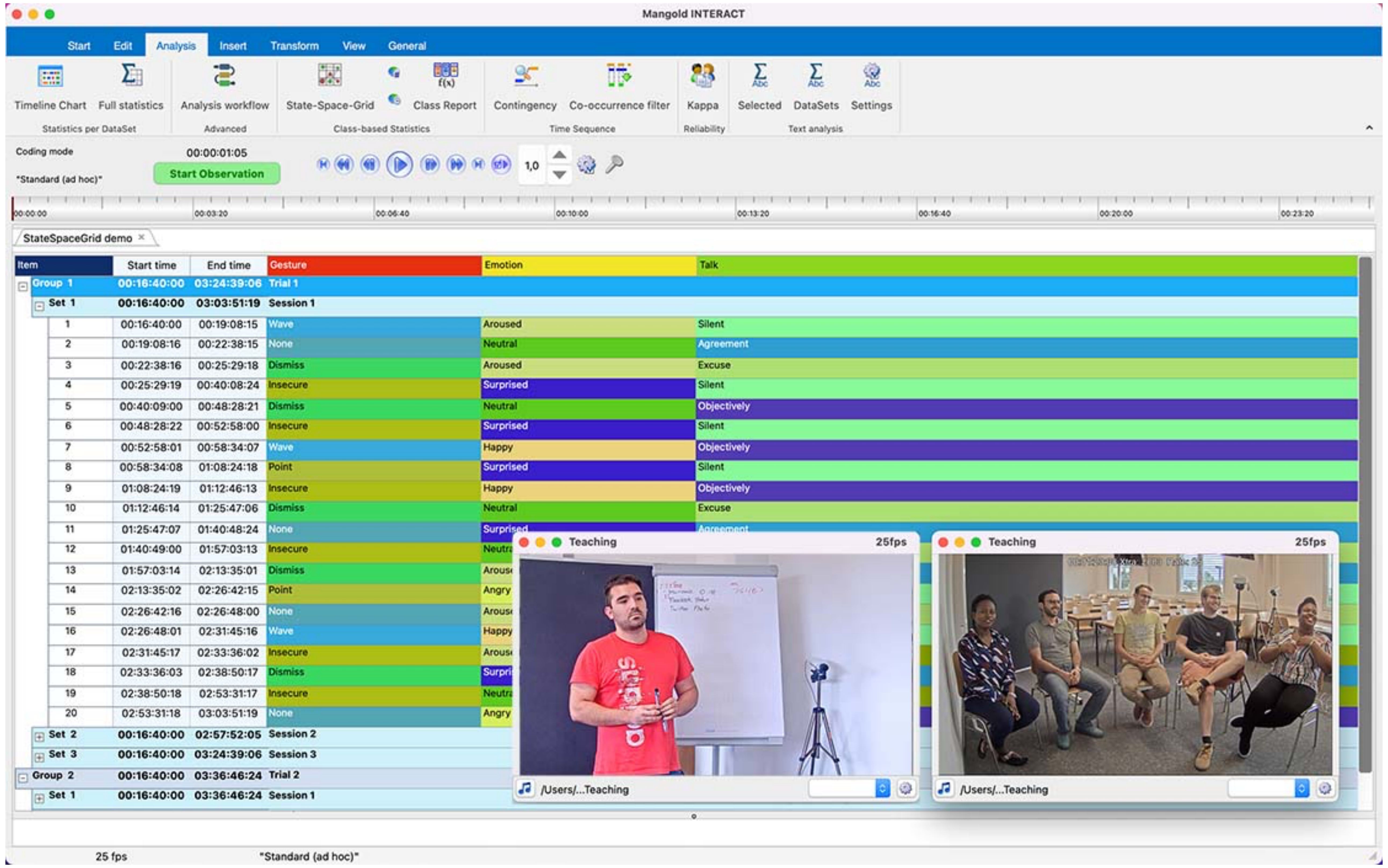
- Image processing - Correction of distortions and deformations (geometric) that occur from camera angles that are not optimal
- Brain imaging - transforming from sensor to brain coordinates, mapping different modalities, standardization for format
- Parallel lines to parallel lines
- e.g. Rotation, Translation, Scaling, Shear
- NiBabel documentation

Integration strategies - Sample time mismatch

- Super/sub sampling with filtering
- Time/sample shift to align data

Integration strategies - Software and format

- What if you have image/video data, EEG, text, audio?
- Each is in its own format, with different sample timings, not keyed events, coordinates, dataframes
- Traditional way?
- Newer way?



Integration strategies - Sample rate variability

- Do you have an accurate time measure and know the variability?
  - Yes - then you can simply interpolate and resample to create a new equally spaced set
- Inaccurate time measure, some information is lost
  - Computer timers for example do not provide accurate time measures unless they are specialized hardware
  - Can assume it's accurate if sampling much much faster than dynamics
  - Reduce sample rate (sub-sample) below estimated variability
  - Cannot use for time-critical associations

Integration strategies- missing data, mixture, non-tabular

- Addressed in earlier lectures (NANs)
- Wrangling
- Manual labor
- May need to use portions of the data
- Large sets need automated or semi-automated detection means

Integration strategies - Memory and processor usage

- Why do we need to be aware of this issue?
- Cloud computing services
- Efficient coding
- Considering data partially, in chunks, computed offline, pre-computed then processed as needed for analysis
- Variable sizes
- n-dimensions - what is the ***curse of dimensionality***?

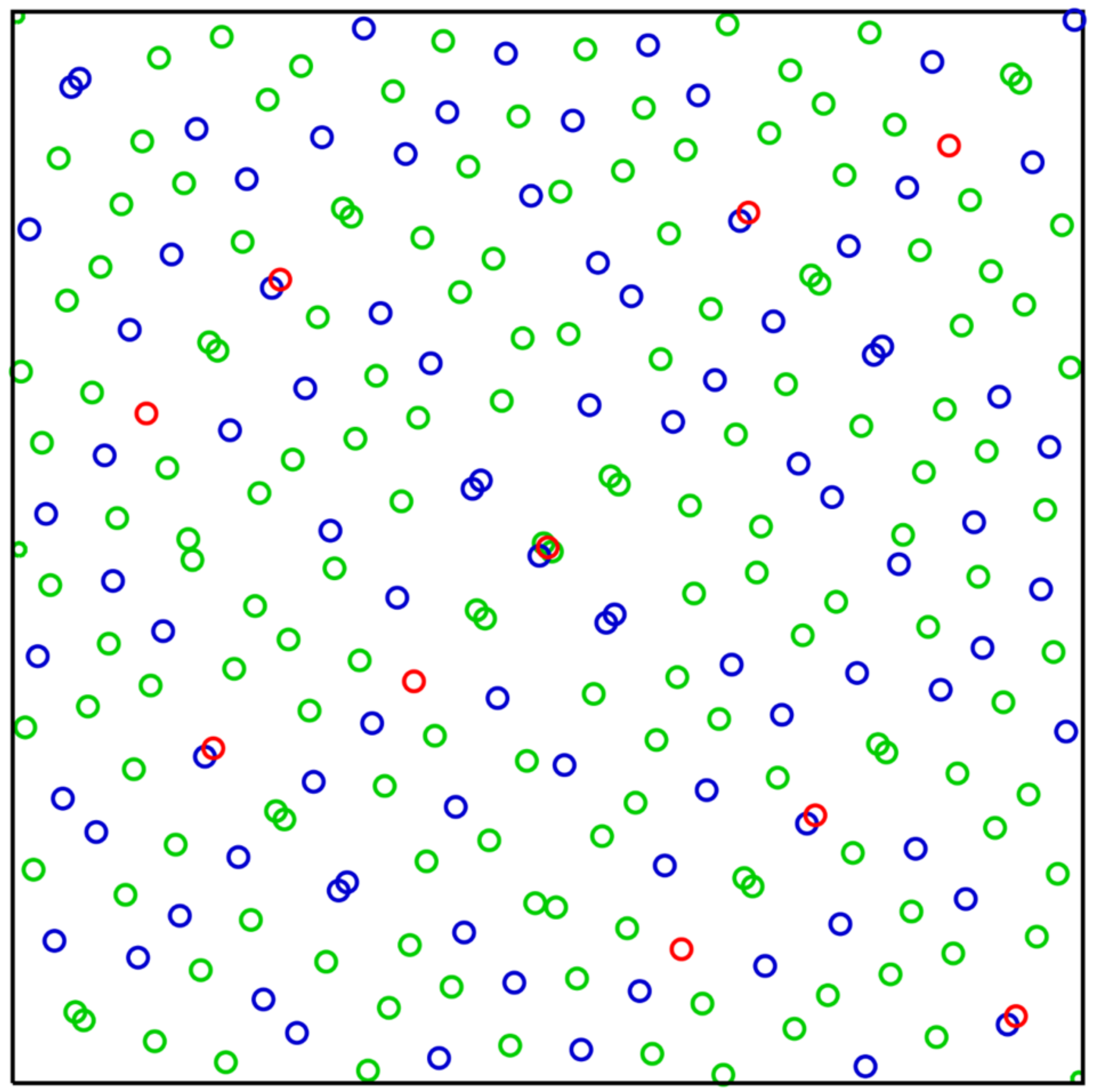
A4: Integrating heterogeneous datasets for neuroscience

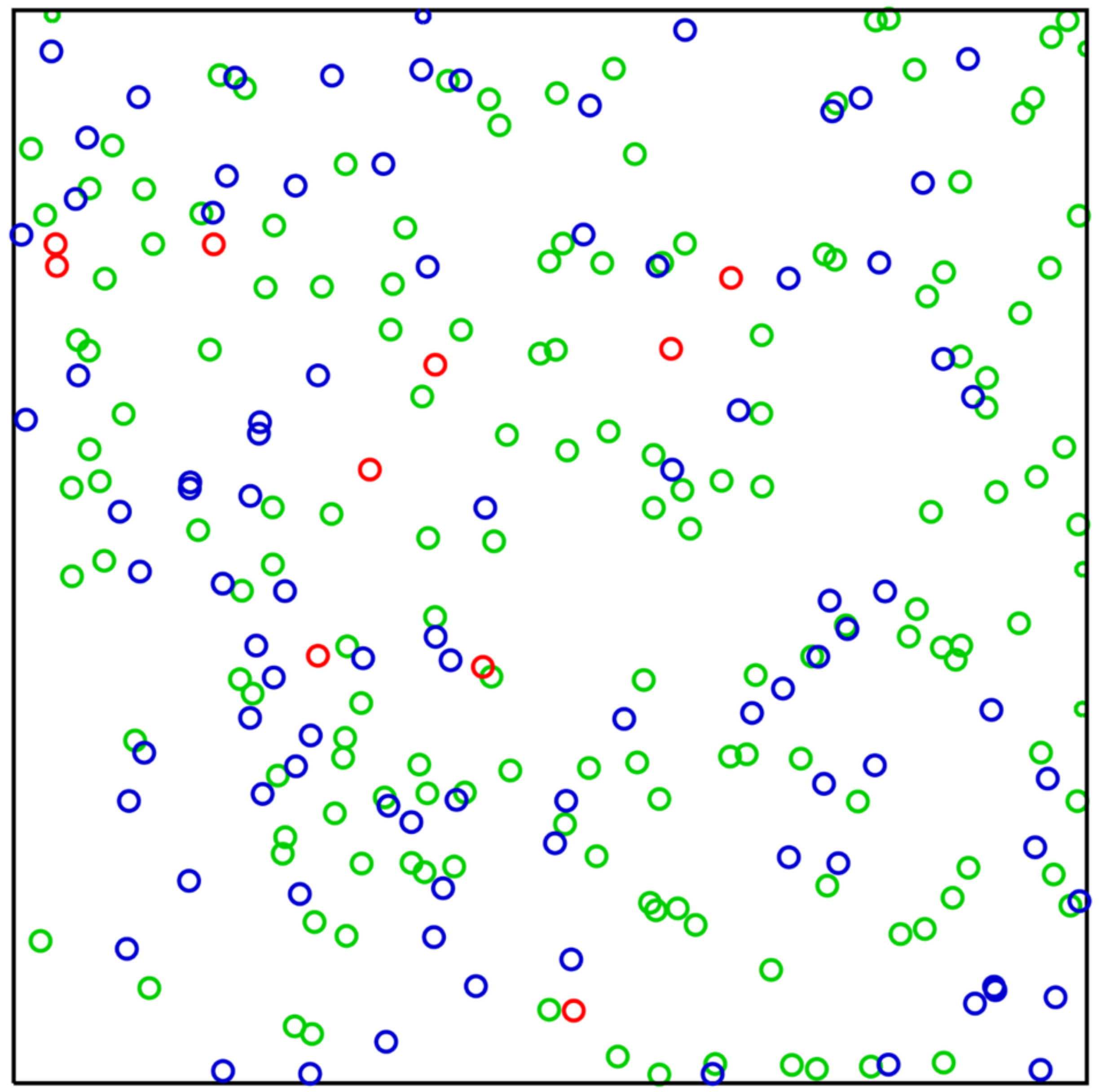
Modules for A4

- **nibabel**
  - Neuroimaging in Python
  - https://nipy.org/nibabel/
- **pysurfer**
  - Visualizing brain imaging data
  - https://pysurfer.github.io
- **sobol_seq**
  - Sobol sequence generator
  - https://github.com/naught101/sobol_seq
  - https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.qmc.Sobol.html

Sobol sequences

- Quasi-random low-discrepancy sequences
- https://en.wikipedia.org/wiki/Sobol_sequence
- Which one covers the space more evenly, just by eye?
  - Sobol or pseudorandom
- **Sobol sensitivity analysis** to analyze influence of parameters in computational neuroscience models
  - https://hal.science/hal-03464025/file/root.pdf
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8184610/
    - Model reproducibility





PySurfer

- PySurfer is a Python library for **visualizing cortical surface representations of neuroimaging data**.
- The package is primarily intended for use with Freesurfer, but it can plot data that are drawn from a variety of sources.
- PySurfer extends Mayavi's powerful rendering engine with a high-level interface for working with MRI and MEG data.

pysurfer - installation

```
pip install pysurfer
```

Dependencies¶

PySurfer works on Python 2.7 and 3.6+. (Older Python 3 versions will probably work, but are not tested.)

To use PySurfer, you will need to have the following Python packages:

- numpy
- scipy
- nibabel
- mayavi
- matplotlib

Some input/output functions also make use of the Python Imaging Library (PIL) and `imageio`, although they are not mandatory.

NiBabel - definition

- "Access a cacophony of neuro-imaging file formats"
- Cacophony?

cacophony | kəˈkäfənē |
noun (pl. **cacophonies**)

a harsh, discordant mixture of sounds: *a cacophony of deafening alarm bells* | *figurative* : *a cacophony of architectural styles* | *songs of unrelieved cacophony.*

- Read and write access to common neuroimaging file formats,
  - including: ANALYZE (plain, SPM99, SPM2 and later), GIFTI, NIfTI1, NIfTI2, CIFTI-2, MINC1, MINC2, AFNI BRIK/HEAD, ECAT and Philips PAR/REC.
  - In addition, NiBabel also supports FreeSurfer's MGH, geometry, annotation and morphometry files,
  - provides some limited support for DICOM

NiBabel - Installation

```
pip install nibabel
```

NiBabel - documentation

- Coordinate systems
- Radiological vs. Neurological conventions
- Intro to DICOM

A4 - Mapping heterogeneous neural data

- How to take different neural data and map them to the human neocortex
- https://en.wikipedia.org/wiki/Human_Connectome_Project
- “A multi-modal parcellation of human cerebral cortex”
  - https://pubmed.ncbi.nlm.nih.gov/27437579/

On to today…

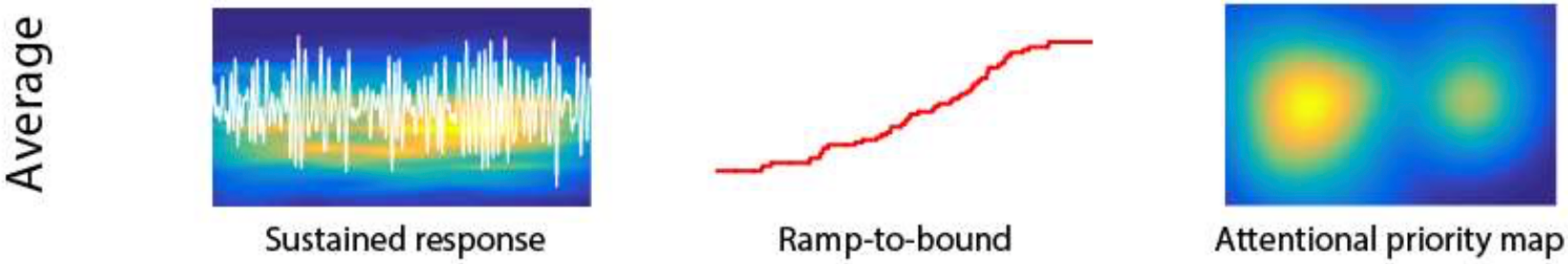
Single trial analysis

Motivation for single trial analysis

- Traditionally neuroimaging techniques are used to compute differences between means over many trials/subjects/studies
  - e.g. in classical cognitive neuroscience, theories of working memory assumed that task-relevant info. is maintained by persistent neural activity
    - Representations kept online by persistent activity patterns, evidence based on averaging massive numbers of subject trials
    - Assumption is if true distribution is contained in noisy measures, measure many times, average, you recover the noise-free representative pattern

So it should look like the following…

- Many single trial measures averaged…
- (Stokes and Spaak 2016)
- Lundqvist,M.etal.(2016) Gamma and Beta Bursts Underlie Working Memory. Neuron



How does the brain really work? Why?

How does the brain really work? Why?

- Brain doesn't operate according to average response
    - Differences in perception, conscious and unconscious processes, real-world embodied, embedded, situated issues (active perception), encoding, decision making
- Strong evidence for high dimensionality of encoding especially in pre-frontal cortex
    - Rigotti, M. et al (2013) The importance of mixed selectivity in complex cognitive tasks. Nature 497, 585–590
- ***Must*** understand neural dynamics within a single trial
- Consider (Stokes and Spaak 2016)

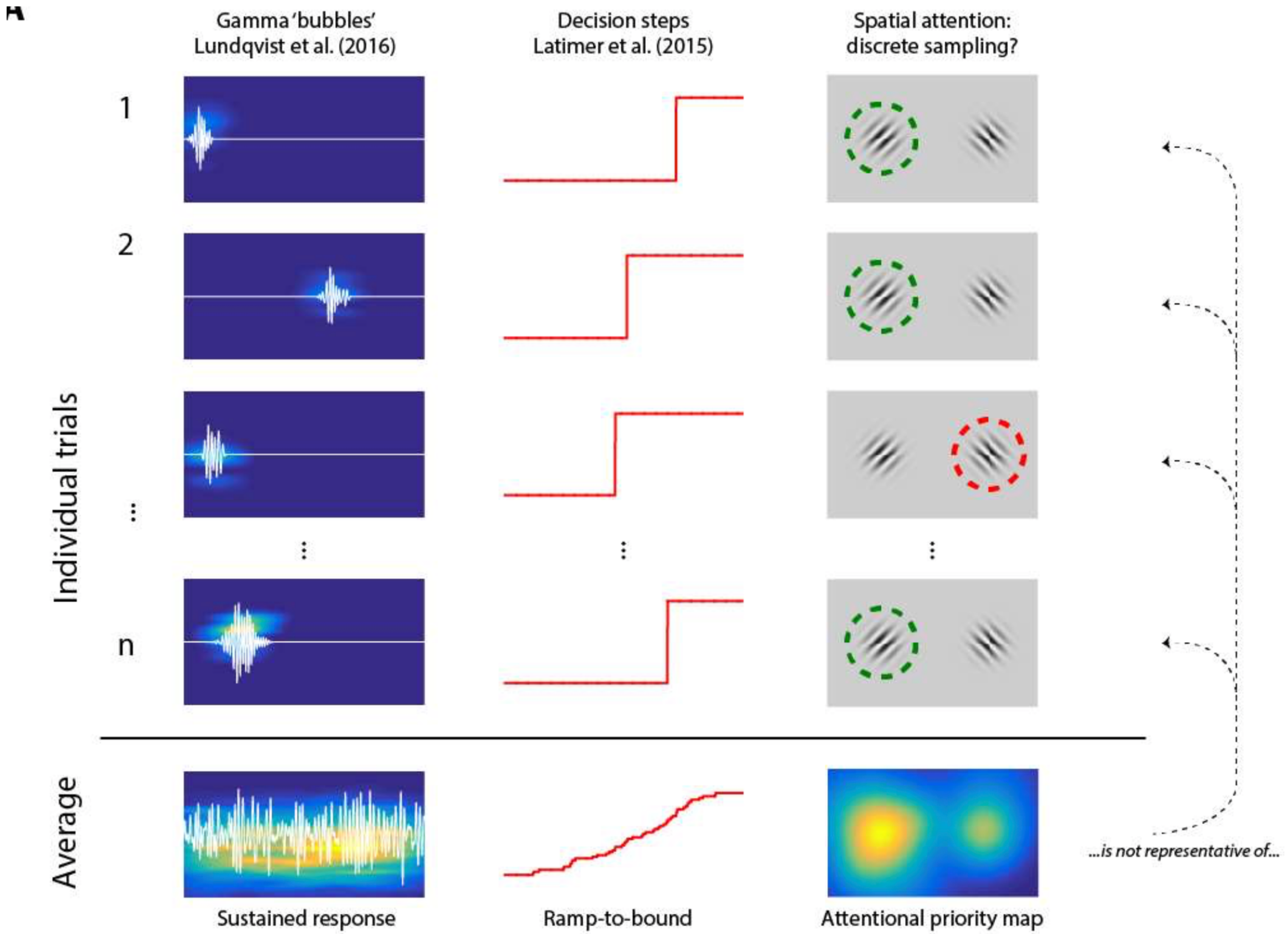
What they tested

- Lunqvist et al. developed novel method to characterize trial-wise dynamics in working memory tasks for primates
- Do we see sustained activity, as previously concluded as recorded from LFP in primate PFC at the single trial?
- Or is it different dynamically at single trial level from ‘average response?’

How did they test it?

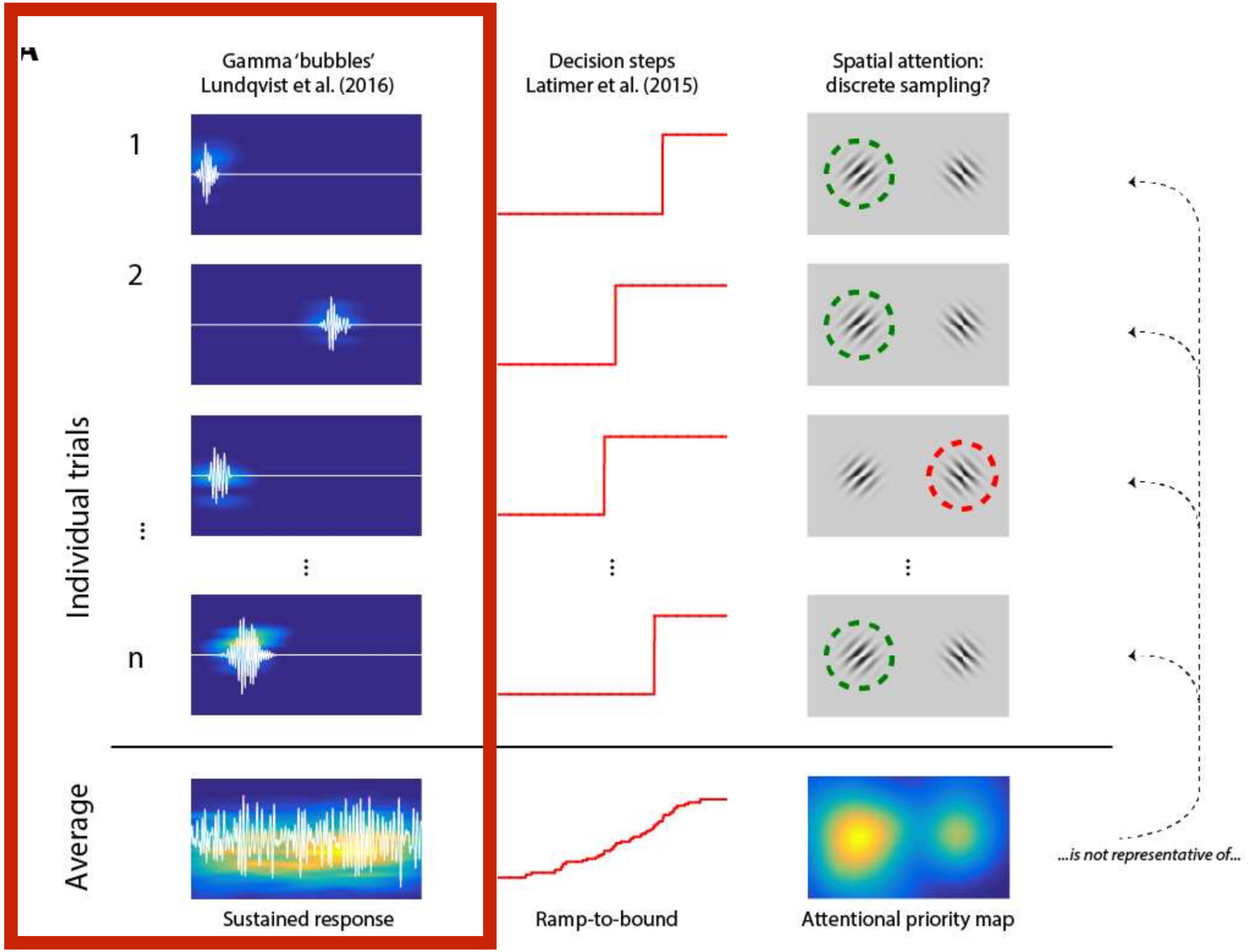
- Developed a novel metric they refer to as ‘burstiness’ to quantify temporal gamma activity for single trials before averaging
- By using 2nd order average of this metric, found persistent activity consists of bursts of activity - not an unbroken chain of firing
- Memories stored in hidden neural states

Combining by computing mean doesn't necessarily create a good representation



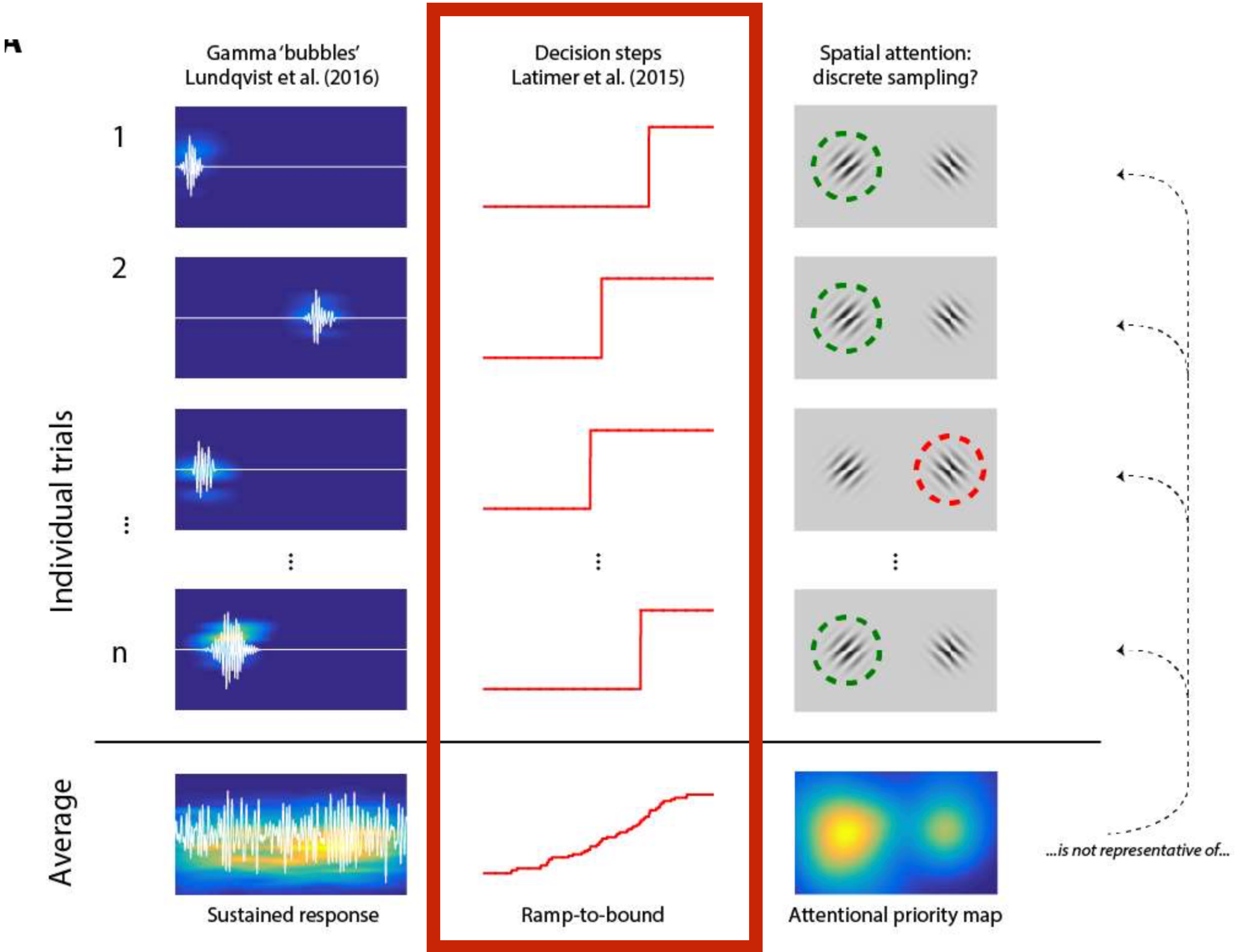
Combining by computing mean doesn't necessarily create a good representation

- Left: (time-frequency representations of power, with traces superimposed) in the prefrontal cortex during individual trials of working memory maintenance activity
- The average shows a familiar sustained gamma response, but qualitatively misrepresents the single trial dynamics.



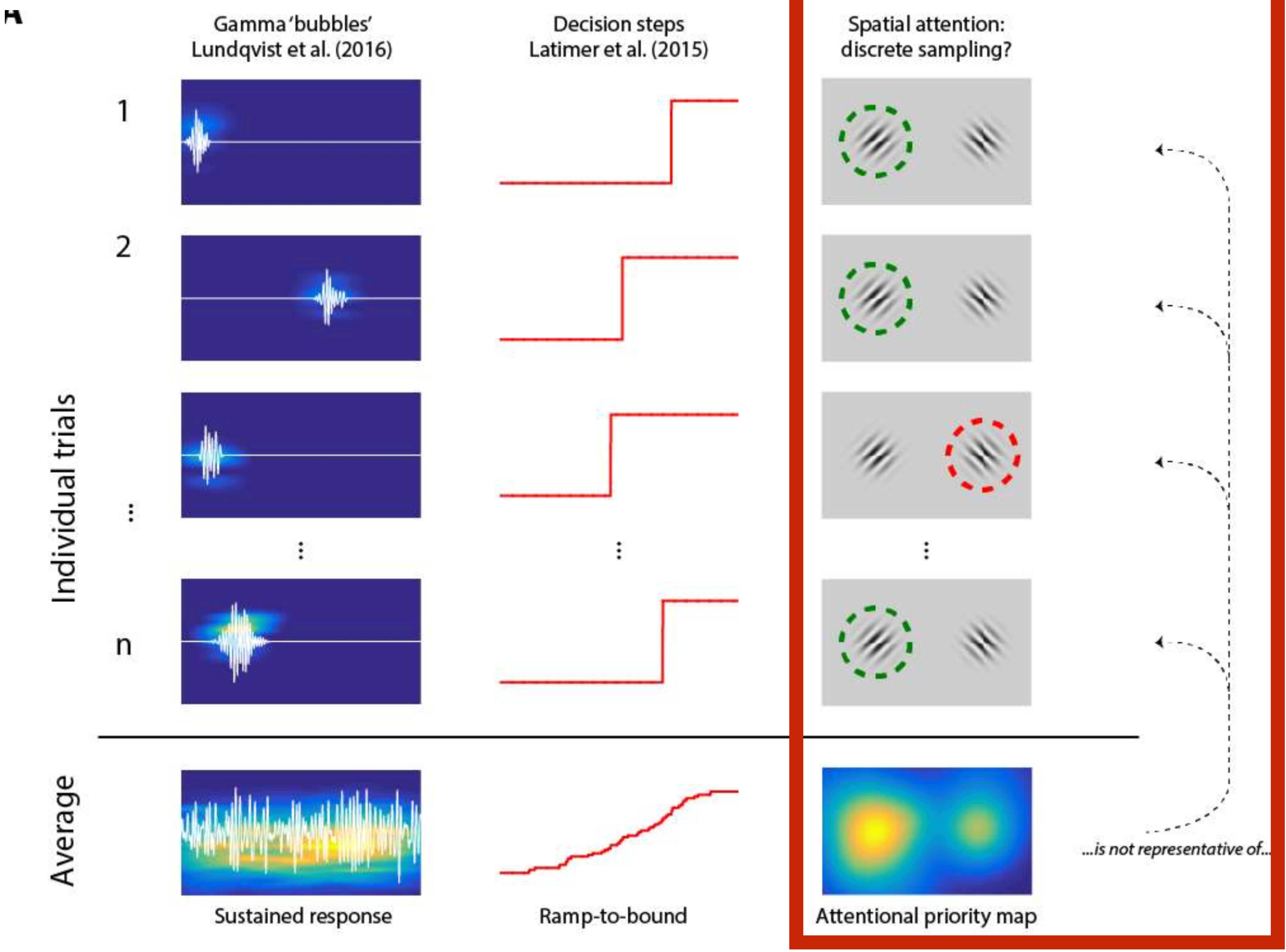
Combining by computing mean doesn't necessarily create a good representation

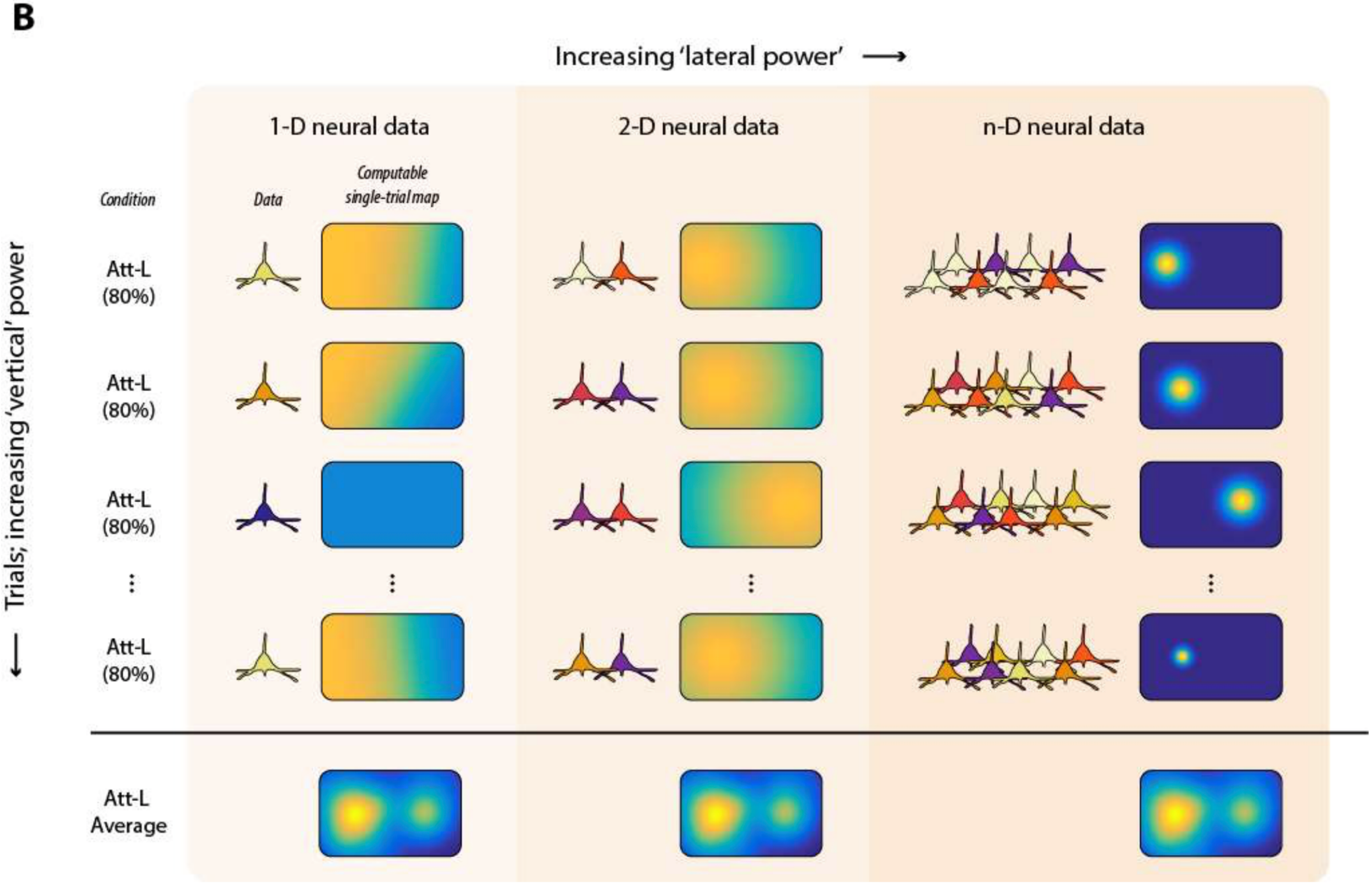
- Middle: neurons display discrete steps reflecting the time of sensory decisions
- The average response shows a classic ramp-to-bound process for the decision.
- Again not representing what's happening



Combining by computing mean doesn't necessarily create a good representation

- Right: spatial attention ***might*** be distributed in a continuous fashion throughout the visual field (as in the average, bottom),
- but such an average profile ***could also be caused by*** individual trials sampling discretely from visual space (80% of trials on the left, 20% on the right).

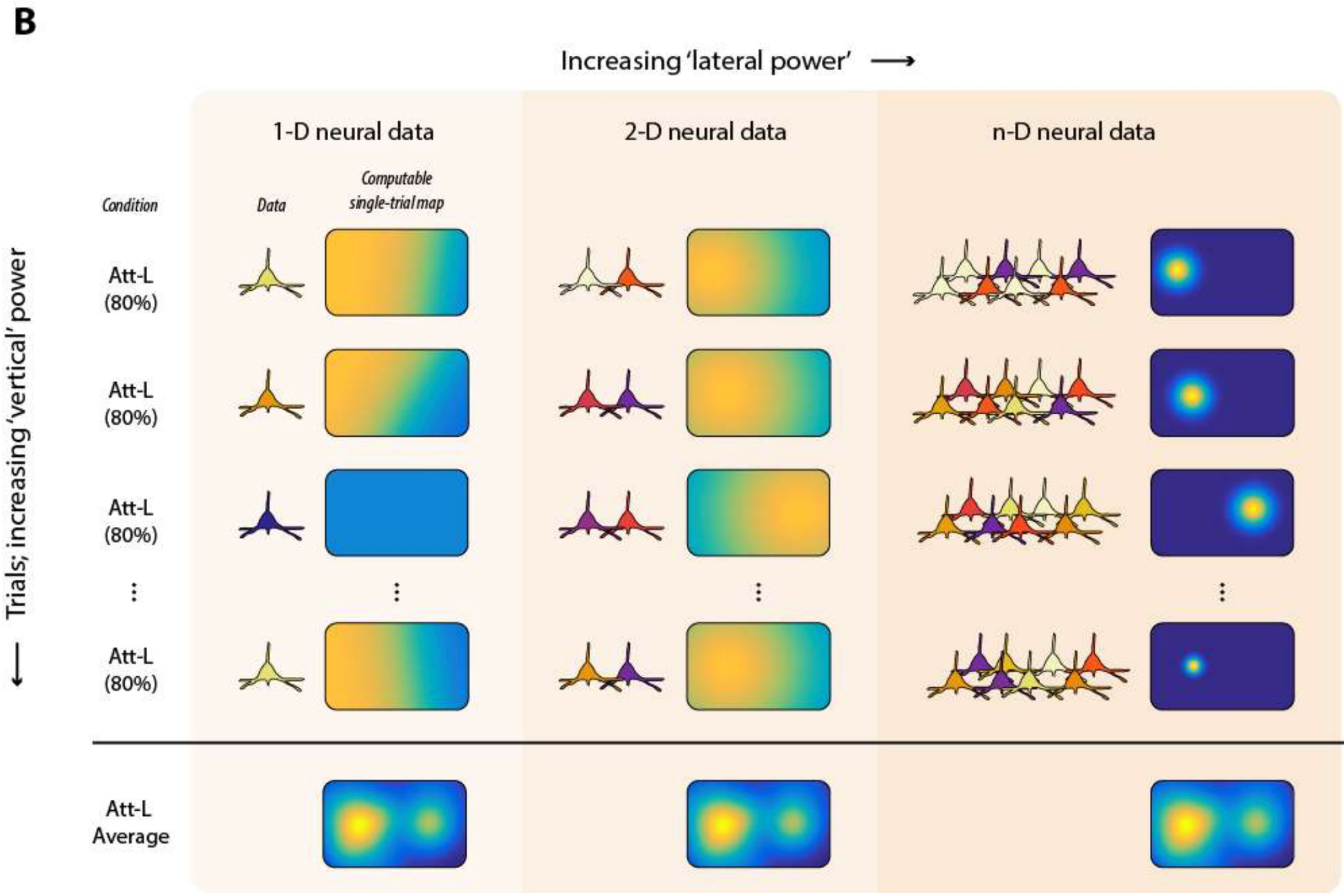



B
Increasing 'lateral power'
1-D neural data
2-D neural data
n-D neural data
Condition
Data
Computable single-trial map
Att-L (80%)
Att-L (80%)
Att-L (80%)
Att-L (80%)
Att-L Average
Trials; increasing 'vertical' power

Combining by computing mean doesn't necessarily create a good representation

- Traditionally **statistical power** = more observations (i.e., trials) to average data
  - "Vertical power"
- **Lateral power:** adding more measurement density (spatial dimension).

Larger lateral power-> Necessary for characterizing neural dynamics in single trial (we'll come back to this)



Motivation for single trial analysis

- We can thus miss important variability patterns by collapsing to the mean
    - Consider a large classroom with 500 students all talking while waiting for lecture.
    - If we record their conversations even for the same class each day all quarter then average the recording, will we recover what the individuals said?
    - No it doesn’t ‘average’ to the conversation as each day has subtle differences
- Solution -**Single trial analysis** studies variability across trials

Single Trial Analysis allows for systematic mapping between

- Brain activity and stimulus information space
- Brain activity and subject behavioral variability
- Brain activity measured using multiple imaging techniques (EEG, fMRI etc)

Single Trial Analysis definition and classes

- All methods that consider ***variance within subjects***
    - 2 classes of methods

        1. Univariate methods

        2. Multivariate methods

- Applications - *Useful for both behavioral and neuroimaging experiments*

STA - Univariate methods

- **Regression over all trials in single subjects measures the relationship between parameterized stimulus space and signal amplitude**
- fMRI - “Parametric design”
- EEG - using same type of approach
  - Neural response to stimulus in individual subjects
  - Probabilistic mapping between stimulus information and EEG amplitude
    - Rousselet G. A., Gaspar C. M., Wieczorek K. P., Pernet C. R. (2011). Modeling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects). *Front. Psychol.* 2:137. 10.3389/fpsyg.2011.00107
  - Time-frequency decomposition of power and phase
    - Cohen M. X., Cavanagh J. F. (2011). Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict. *Front. Psychol.* 2:30. 10.3389/fpsyg.2011.00030

STA - Univariate methods

- **Variance among trials contains info regarding subjects and their cognitive states**

  - e.g. study establishing increased variance in latency of P1 response to Gabor patches than controls for children with autism
    - Milne E. (2011). Increased intra-participant variability in children with autistic spectrum disorders: evidence from single-trial analysis of evoked EEG. *Front. Psychol.* 2:51. 10.3389/fpsyg.2011.00051

  - e.g. pre-stimulus alpha power correlated with subject judgment of state of attention
    - Macdonald J. S. P., Mathan S., Yeung N. (2011). Trial-by-trial variations in subjective attentional state are reflected in ongoing prestimulus EEG alpha oscillations. *Front. Psychol.* 2:82. 10.3389/fpsyg.2011.00082
    - *(review)* VanRullen R., Busch N. A., Drewes J., Dubois J. (2011). Ongoing EEG phase as a trial-by-trial predictor of perceptual and attentional variability. *Front. Psychol.* 2:60. 10.3389/fpsyg.2011.00060

STA - Multivariate methods

- **Often derive pattern classifiers to characterize the spatial-temporal variance in each trial**
- e.g. Touryan J., Gibson L., Horne J. H., Weber P. (2011). Real-time measurement of face recognition in rapid serial visual presentation. Front. Psychol. 2:42. 10.3389/fpsyg.2011.00042
    - Used variance in time/space to train a discriminant function that could classify brain activity associated with familiar/unfamiliar faces in real-time
    - Group ERPs could be used to differentiate over frontal AND parietal electrodes, but with the above methods, ONLY parietal response allowed categorical discrimination on single-trial basis
    - So averaging can actually create misleading illusory signals that are not actually present in individual subjects!
        - Gaspar C. M., Rousselet G. A., Pernet C. R. (2011). Reliability of ERP and single-trial analyses. *Neuroimage* 58, 620–629 10.1016/j.neuroimage.2011.06.052

STA - An additional dimension

- Allows for interpretation of individual differences to quantify effects within and between subjects

- An additional window into brain function

- Rich data description can help expose subtle brain mechanisms that may be hidden when looking at traditionally pooled data (averaged)

STA - caveats -> Requirements

- Many trials (“Vertical power”)
  - To reduce Signal to Noise Ratio - regression over trials
  - Have to be careful not to smooth over important heterogeneity
  - Metrics were developed of ‘burstiness’) - needs a priori model
- Dense coverage (“Lateral power”)
  - For good patterns - time/frequency intervals, localization, avoiding missing spikes in activity, sparse behaviors, etc
- e.g. Rousselet G. A., Husk J. S., Bennett P. J., Sekuler A. B. (2008). Time course and robustness of ERP object and face differences. *J. Vis.* 8, 3, 1–18 10.1167/8.12.3

STA - toolboxes

- **Recipes**
  - Parra L. C., Spence C. D., Gerson A. D., Sajda P. (2005). Recipes for the linear analysis of EEG. *Neuroimage* 28, 326–341 10.1016/j.neuroimage.2005.05.032
- **PyMVPA (http://www.pymvpa.org)**
  - Hanke M., Halchenko Y. O., Sederberg P. B., Hanson S. J., Haxby J. V., Pollmann S. (2009). PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7, 37–53 10.1007/s12021-008-9041-y
- **EEGLAB, SIFT, NFT, BCILAB, and ERICA**
  - Delorme A., Mullen T., Kothe C., AkalinAcar Z., Bigdely-Shamlo N., Vankov A., Makeig S. (2011). EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Comput. Intell. Neurosci.* 2011, 130714.

STA - more toolboxes

- Hartmann T., Schulz H., Weisz N. (2011). Probing of brain states in real-time: introducing the console environment. *Front. Psychol.* 2:36. 10.3389/fpsyg.2011.00036
- FieldTrip
    - Oostenveld R., Fries P., Maris E., Schoffelen J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell Neurosci.* 2011, 156869.
- Pernet C. R., Chauveau N., Gaspar C., Rousselet G. A. (2011). LIMO EEG: a toolbox for hierarchical linearmodeling of electroencephalographiuc data. *Comput. Intell. Neurosci.* 2011, 831409.

In class report development (~30m)

- Define this course’s intent
- Draw comparisons between this course and requirements
- How does this course build upon what came before?
- How can you use your starting point in this course to expand your understanding?