

Information theoretic analysis of EEG using Gaussian-Copula Mutual Information (GCMi)

Robin Ince



University
of Glasgow

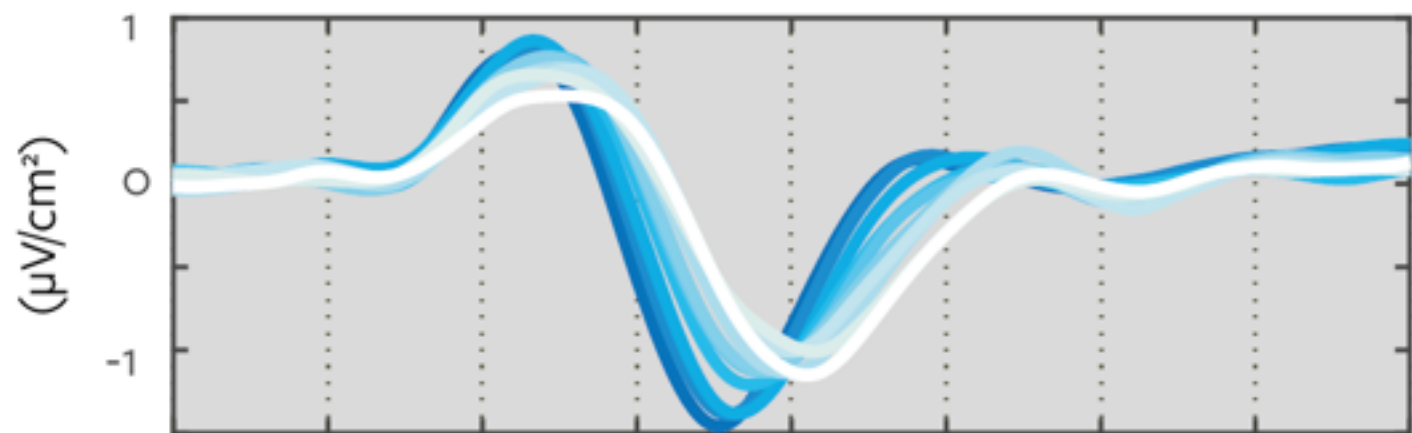
Outline

- Introduction to GCMI
- Practical 1: Two category event related EEG
- Practical 2: Continuous Feature - Bubbles Sampling
- Practical 3: Representational interactions

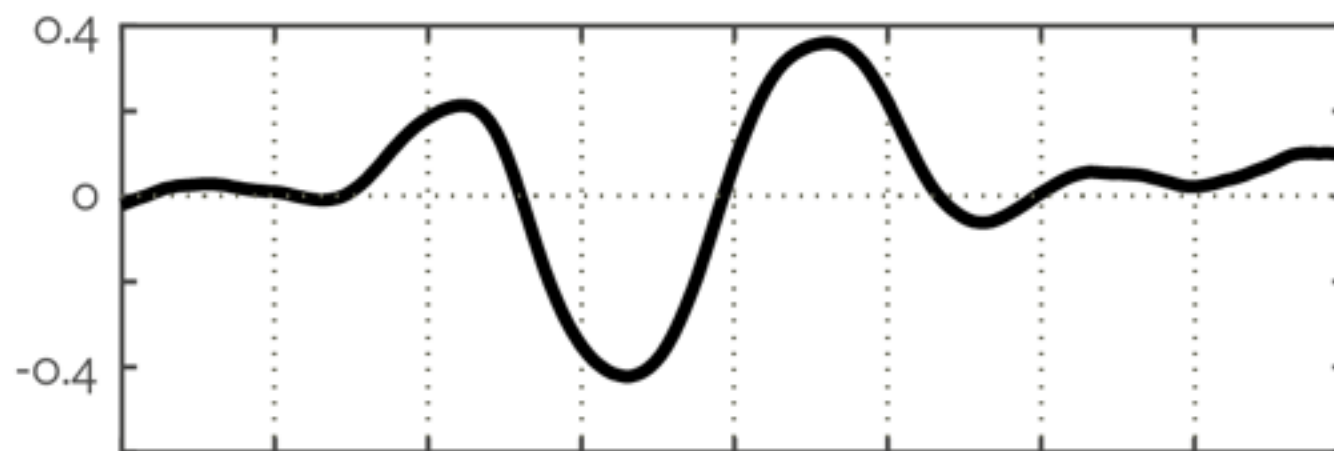
Neuroimaging Data Analysis

- Goal of most (functional) neuroimaging data analysis (within cognitive neuroscience): Detect and quantify modulations of recorded signals [EEG, MEG, fMRI] by experimental stimuli or conditions
- Statistics:
 - Determine “statistical significance” (reject null hypothesis of no effect)
 - Measure size of the effect
 - Eg: d' , t-test, correlation, ANOVA, GLM, multivariate decoding approaches

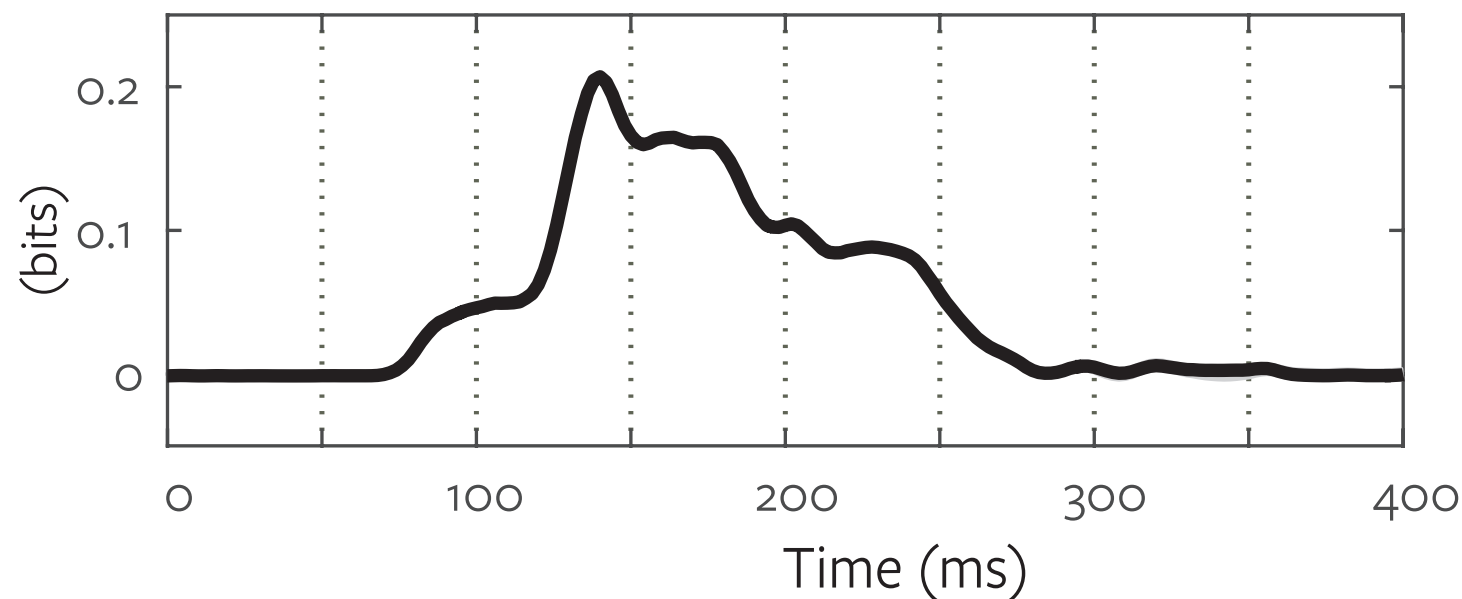
Where and how strongly does the stimulus affect the recorded signal?



ERP
(eye deciles)

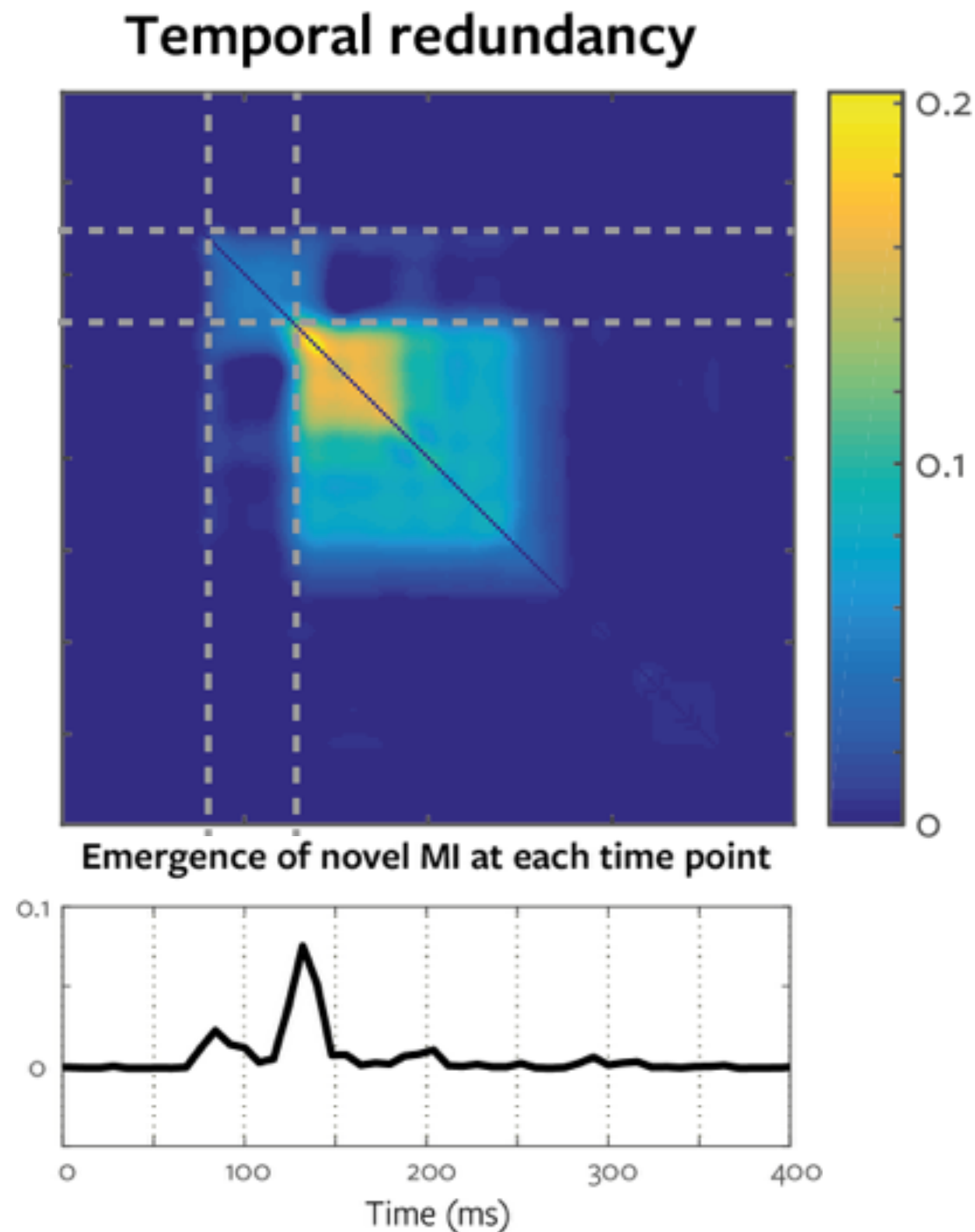


Rank Correlation
(eye, EEG)



Mutual Information

How does the stimulus representation emerge over time?



Mutual Information

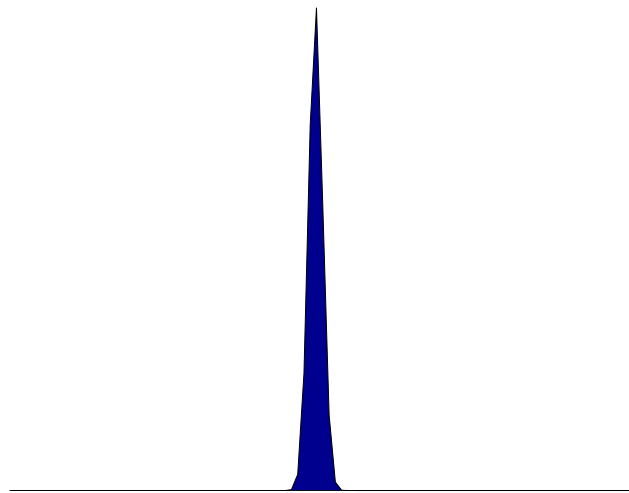
- Mutual Information (MI) is the effect size for a statistical test of dependence (against null hypothesis that the two variables are statistically independent)
- Most general assumption/model free such test (not restricted to linear effects)
- Difficult to estimate in practise
- Nice interpretations: coding/decoding, ideal observer, yes-no questions, average single trial reduction in uncertainty

Entropy

- MI is based on **entropy**; a measure of uncertainty (cf. variance)
- High entropy = high variance; low entropy = low variance; but entropy not restricted to unimodal variables
- Many information theoretic quantities have analogues in traditional statistics - simply replace variance with entropy (eg ANOVA - MI)



High entropy, High variance



Low entropy, low variance



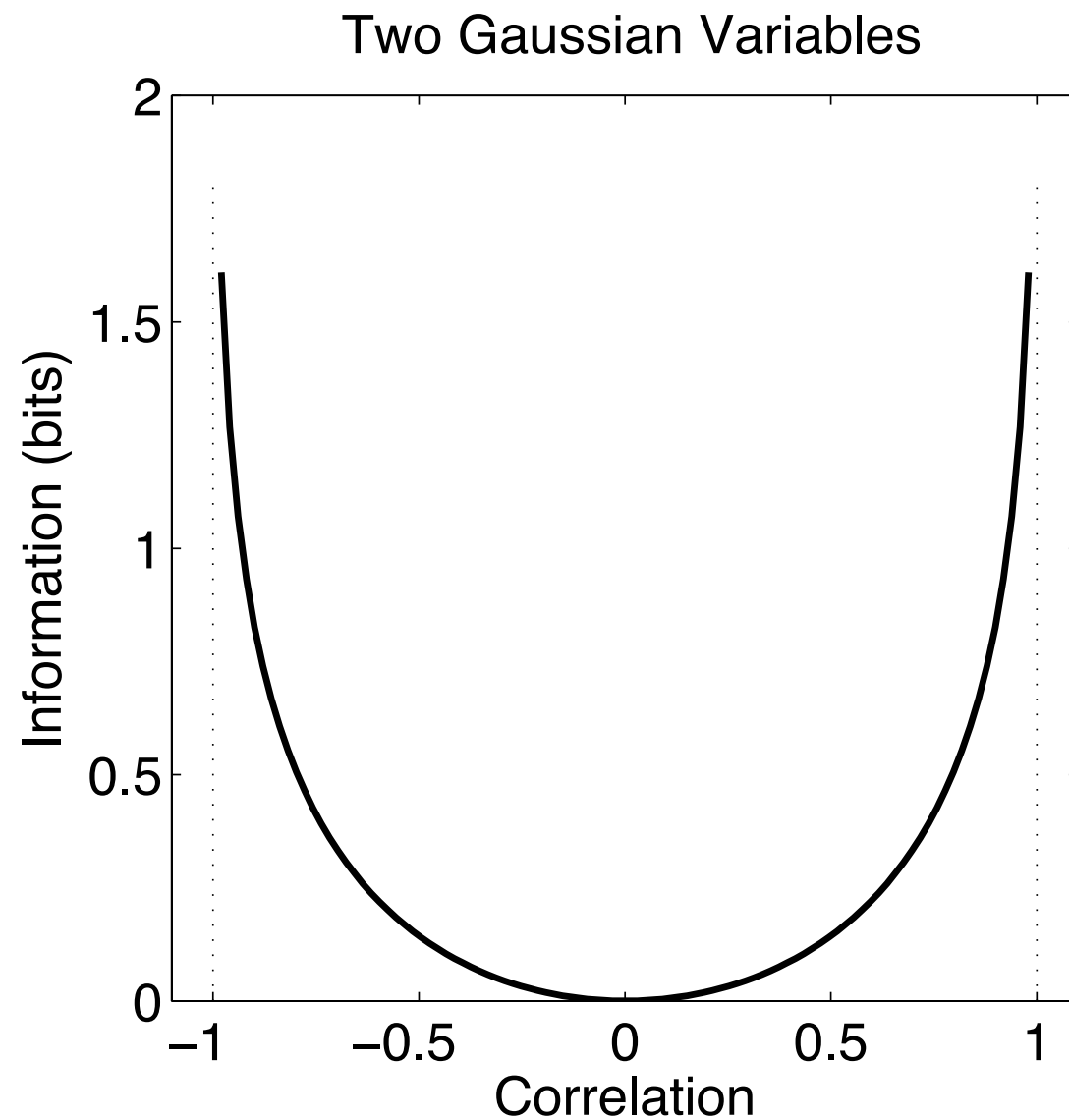
Low entropy, high variance

Mutual Information

$$\begin{aligned} I(R; S) &= H(R) - H(S) - H(R, S) \\ &= H(R) - H(R|S) \\ &= H(S) - H(S|R) \end{aligned}$$

- 3 forms - each lead to an interpretation
- Replace variance with entropy - gives information theoretic analogues to common statistics (think of variance explained)
- Meaningful effect size - units of bits

Mutual Information



- Unsigned; higher contrast than correlation

Calculating Mutual Information

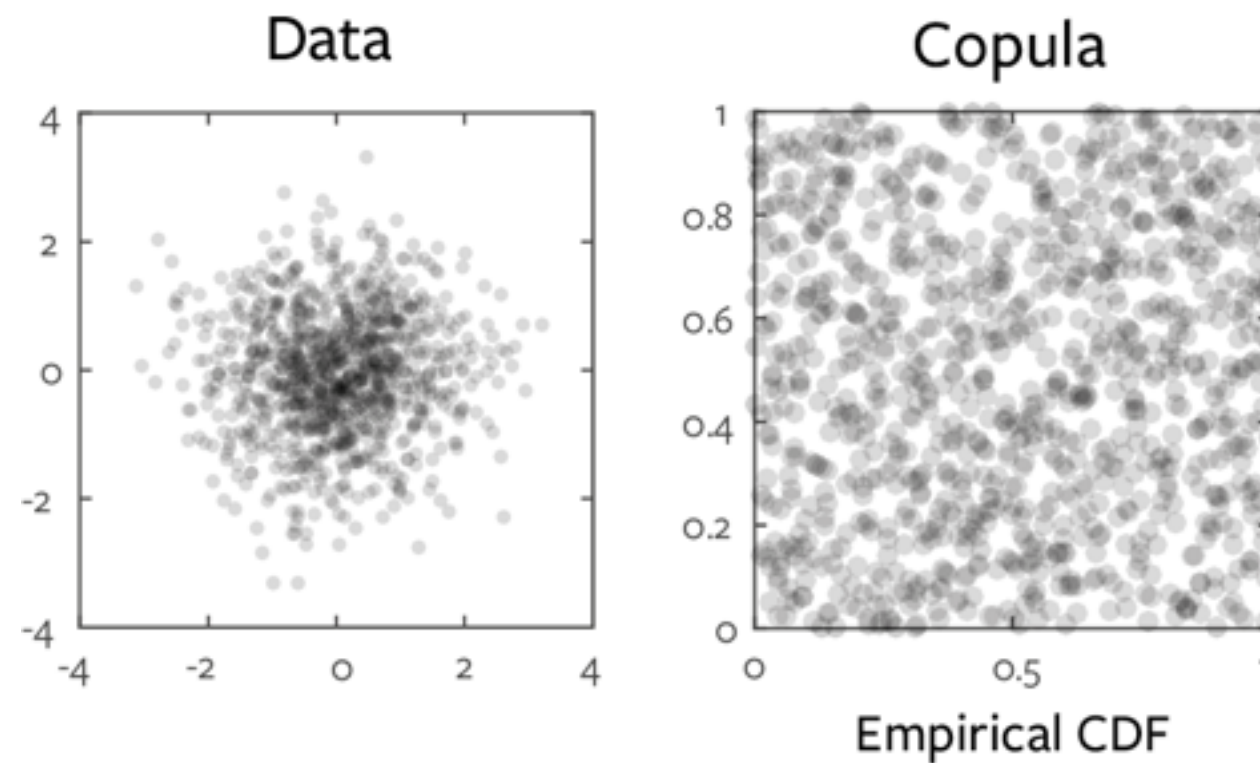
- Discrete formulation. Bin continuous data. Bias, infeasible for multi-dimensional signals.
- Nearest Neighbour (Kraskov). Low bias, but very high variance. Computationally expensive.
- Parametric. (i.e. assume data are Gaussian). Great if assumptions are met.
- Gaussian-Copula Mutual Information. Computationally cheap. Robust. Data efficient. BUT an approximation

GCMi

- <https://github.com/robince/gcmi>
- <http://onlinelibrary.wiley.com/doi/10.1002/hbm.23471/full>
- Works by “normalizing” data and then applying a Gaussian assumption on the dependence (individual input variables do not have to be Gaussian!)
- Mathematically justified (lower bound on true information)
- Rank statistic (input is ranks of each variable)

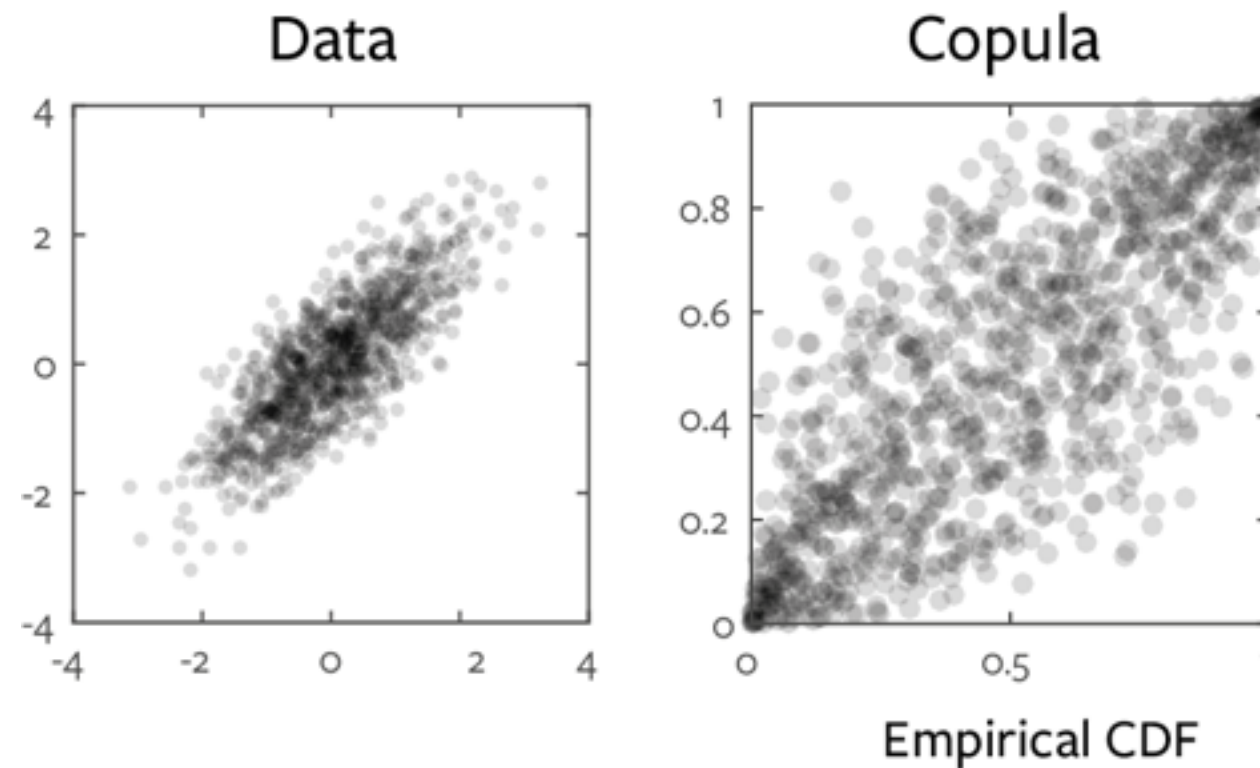
GCMI

- “copula” = maths-y way of saying we look at ranks (ignore the marginal distribution)



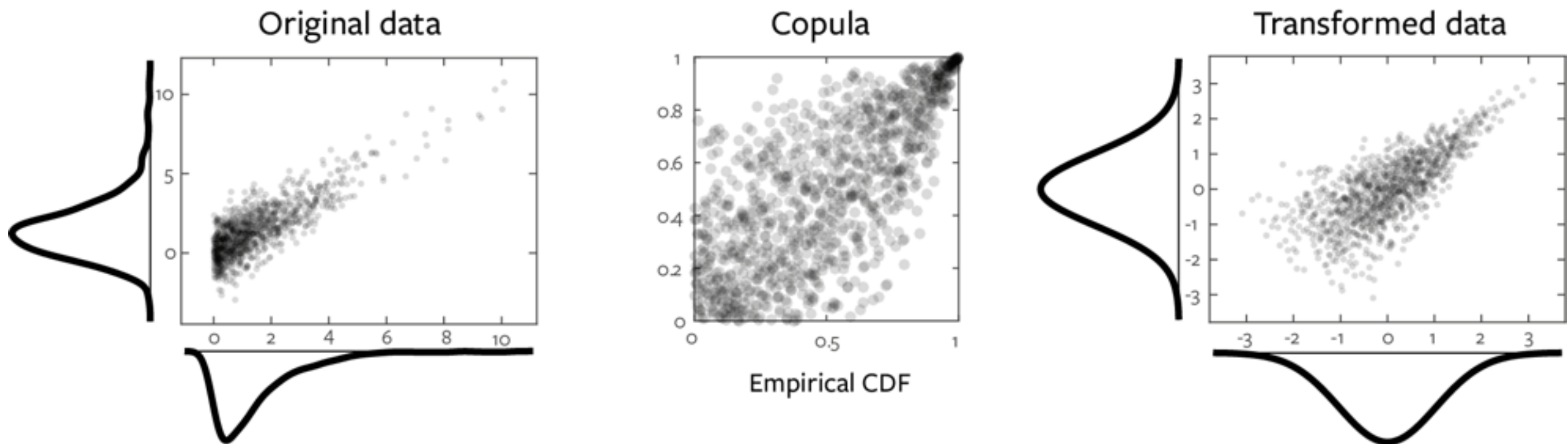
GCM

- “copula” = maths-y way of saying we look at ranks (ignore the marginal distribution)

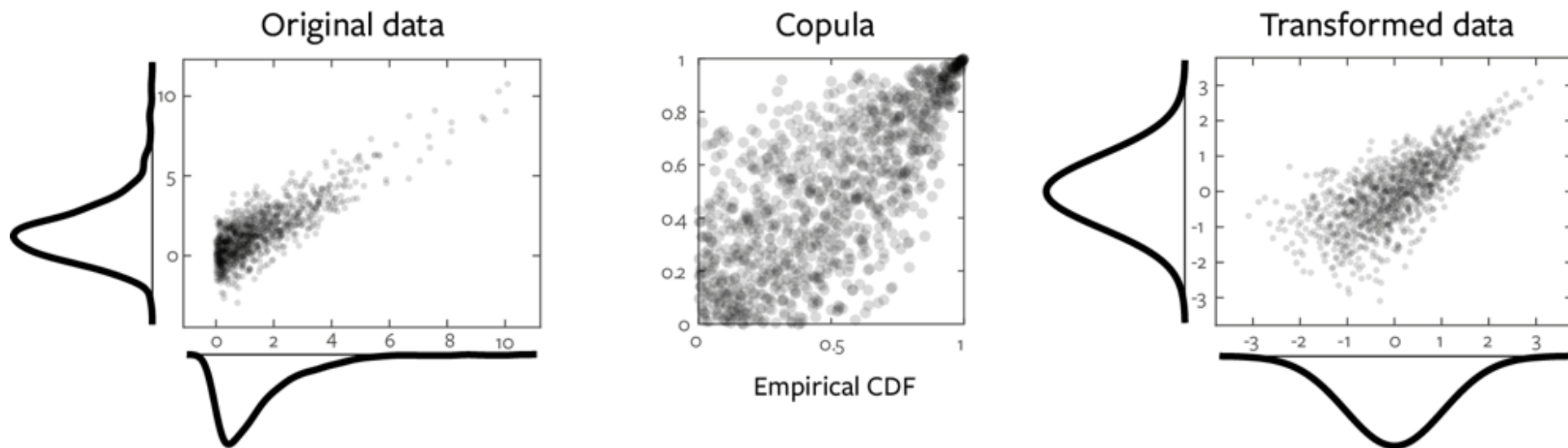


GCM

- Transform marginals to standard normal preserving rank relationships



GCMI



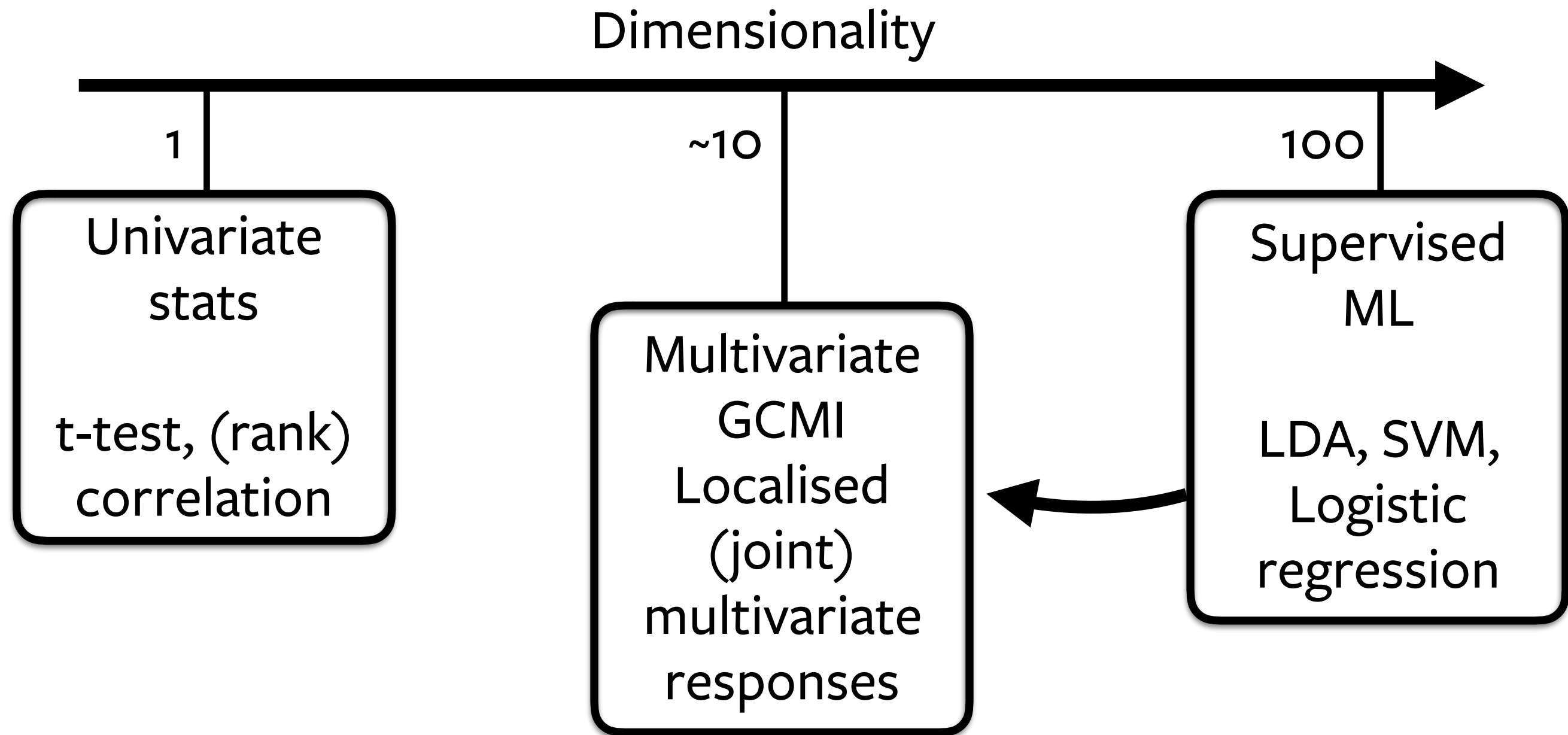
- Transform marginals to standard normal preserving rank relationships (`copnorm` function)
- Use Gaussian parametric estimation (`mi_gg` function)
- Gives a lower bound MI estimate (`gcmi_cc` does both steps)

GCMi

- Multivariate
- Rank based (robust)
- Effect sizes on a meaningful, additive common scale
- Can combine discrete and continuous variables
- Equivalent statistical power to existing methods (e.g. t-test, rank correlation, etc.)
- Use with permutation testing
- Easy to use

```
r = corr(X, Y);  
I = gcmi_cc(X, Y);
```

Multivariate Responses



```
I_A = gcmi_cc(RA, Stim);  
I_B = gcmi_cc(RB, Stim);  
I_jointAB = gcmi_cc([RA RB]; Stim);
```

Multivariate MI

- For multidimensional variables, copula transform each dimension independently
- Can apply to low dimensional multivariate responses (1-10 dim)
e.g. magnetic field vectors, EEG voltage + instantaneous temporal derivative, complex spectra
- Allows for higher-order information theoretic quantities :
conditional mutual information, interaction information,
directed information (transfer entropy), directed feature
information

MI as the basis of framework for data analysis

- Robust (rank based), computationally efficient, meaningful effect size (bits), common scale (across univariate, multivariate, continuous and discrete response variables, behaviour etc.)
- **Conditional Mutual Information** - (like partial correlation) condition out the effect of correlated features (also group statistics)
- **Interaction Information** - study representational interactions (c.f. RSA, temporal generalisation decoding)
- **Directed Information** (transfer entropy), **Directed Feature Information** (communication of specific content) (Ince et al. *Scientific Reports* 2015)

Information theoretic quantity	Other statistical approaches
Mutual Information (discrete; discrete)	Chi-square test of independence; Fishers exact test
MI (univariate continuous; discrete)	2 classes: T-test, KS-test, Mann-Whitney U test; ANOVA
MI (multivariate continuous; discrete)	2 classes: Hotelling T ² -test; Decoding (CV classifier)
MI (univariate continuous; univariate continuous)	Pearson correlation; Spearman rank correlation; Kendall rank correlation
MI (multivariate continuous; univariate continuous)	Generalized Linear Model framework Decoding (CV regression)
MI (multivariate continuous; multivariate continuous)	Canonical correlation analysis Distance correlation
Conditional Mutual Information	Partial correlation (continuous variables and linear effects only)
Directed Information	Granger causality
Directed Feature Information	Dynamic Causal Modeling (Psychophysiological Interactions)
Interaction Information	Representational Similarity Analysis (redundancy only) Cross-classification decoding (redundancy only) Mediation analysis

Information theoretic quantity

Mutual Information (discrete; discrete)

MI (univariate continuous; discrete)

MI (multivariate continuous; discrete)

MI (univariate continuous; univariate continuous)

MI (multivariate continuous; univariate continuous)

MI (multivariate continuous; multivariate continuous)

Conditional Mutual Information

Directed Information

Directed Feature Information

Interaction Information

Other statistical approaches

Chi-square test of independence; Fishers exact test

2 classes: T-test, KS-test, Mann-Whitney U test; ANOVA

2 classes: Hotelling T²-test; Discriminant Analysis (CV classifier)

Pearson correlation; Spearman rank correlation; Kendall rank correlation

Generalized Linear Model framework
Decoding (CV regression)

Canonical correlation analysis
Distance correlation

Partial correlation (continuous variables and linear effects only)

Granger causality

Dynamic Causal Modeling (Psychophysiological Interactions)

Representational Similarity Analysis (redundancy only)
Cross-classification decoding (redundancy only)
Mediation analysis

**Common,
quantitatively
comparable
effect size**



Questions?

Practical 1

Two category event related EEG

prac1_discrete_eeg.m

Data

- Face perception EEG data set (thanks to Guillaume Rousselet, Kasia Jaworska)
- 2 classes: Face (stim=0) vs Noise (stim=1)
- CSD preprocessing
- `csddat` : `[Ntr1 Nch Nt]`
`stim` : `[1 Ntr1]`
`time` : `[1 Nt]`
`chanlocs`: `[1 Nch]`

GCM - Continuous-Discrete

- Two different approaches
- `mi_model_gd` : a model comparison more like ANOVA. Compares a rank-Gaussian unconditional model, to conditional rank-Gaussian models. A lower bound in 1D, but not in 2D+.
- `mi_mixture_gd` : estimates MI based on a mixture of approximate rank-Gaussian conditional.

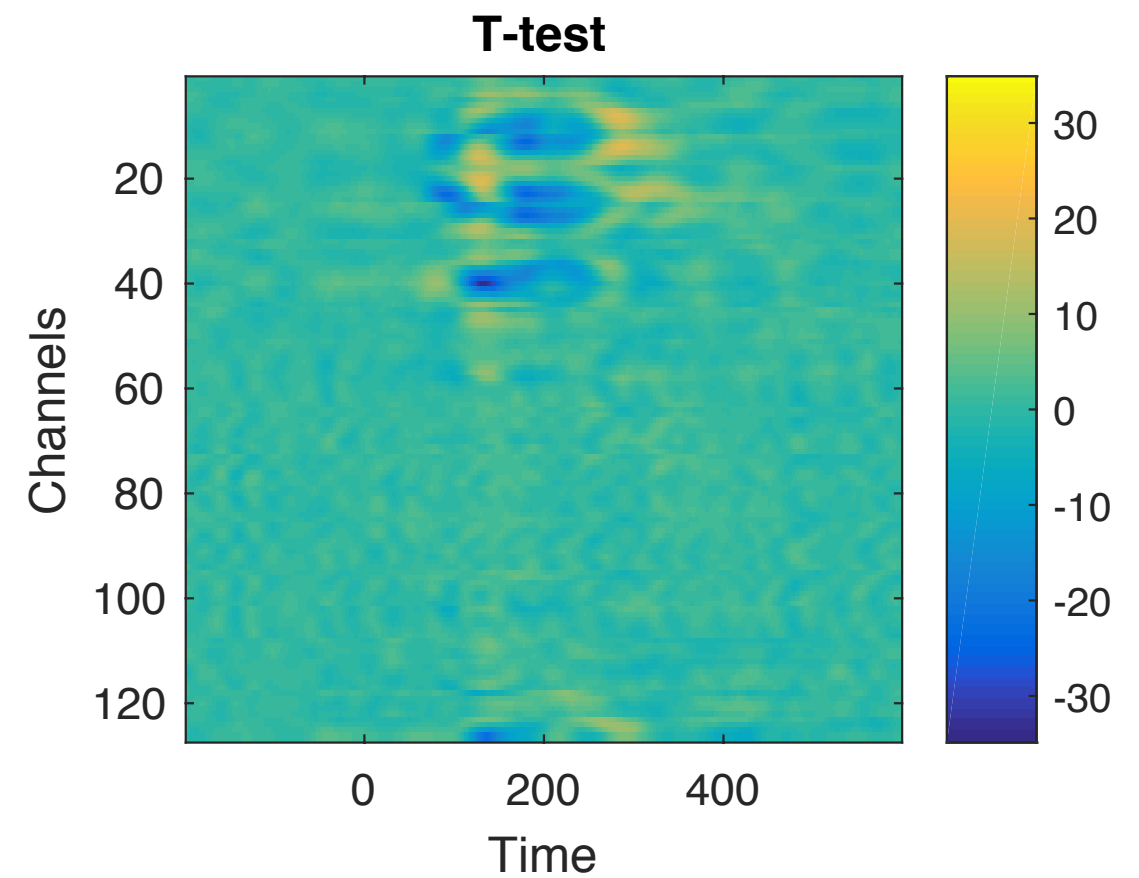
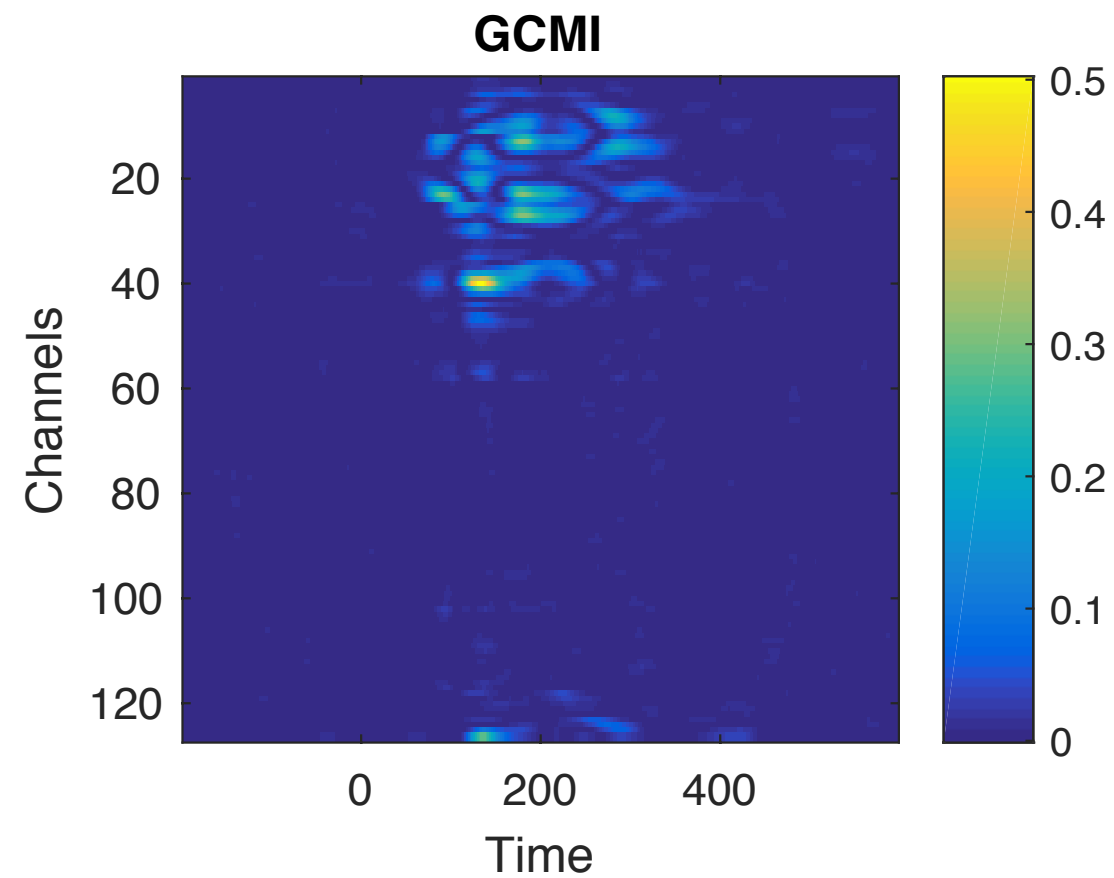
GCM1 - Continuous-Discrete

- `mi_model_gd` : better statistical power, computationally faster. Use by default when you are doing conventional statistics.
- `mi_mixture_gd` : Use for higher dimensional responses when you plan to do something quantitative with the resulting MI values (e.g. comparing with behaviour, calculating interaction information).

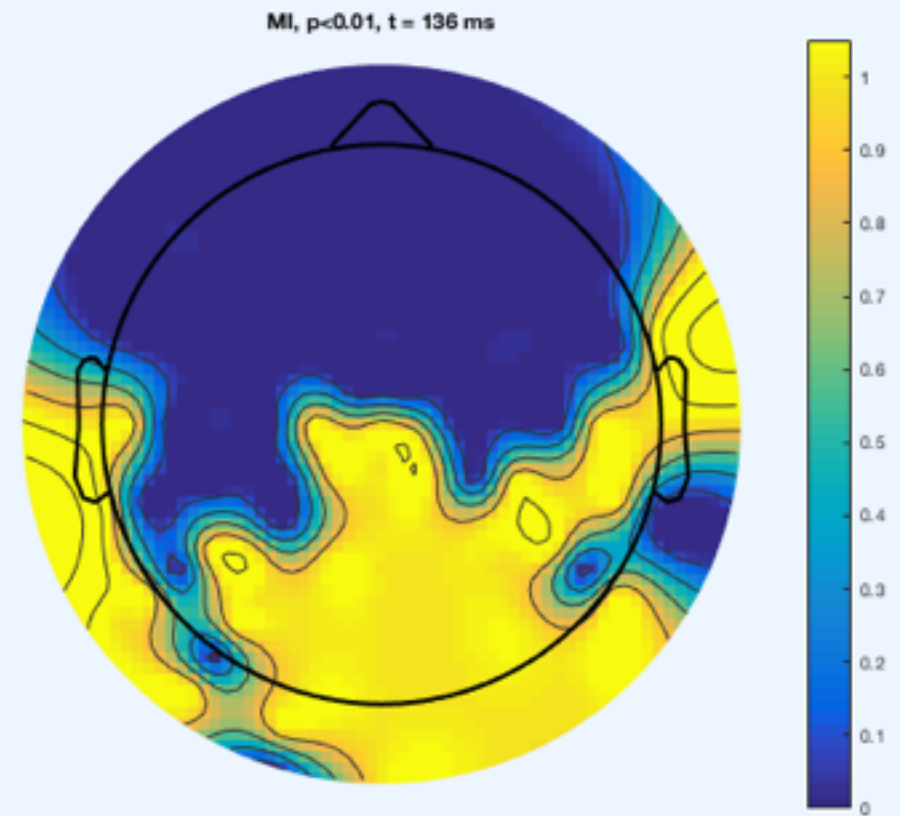
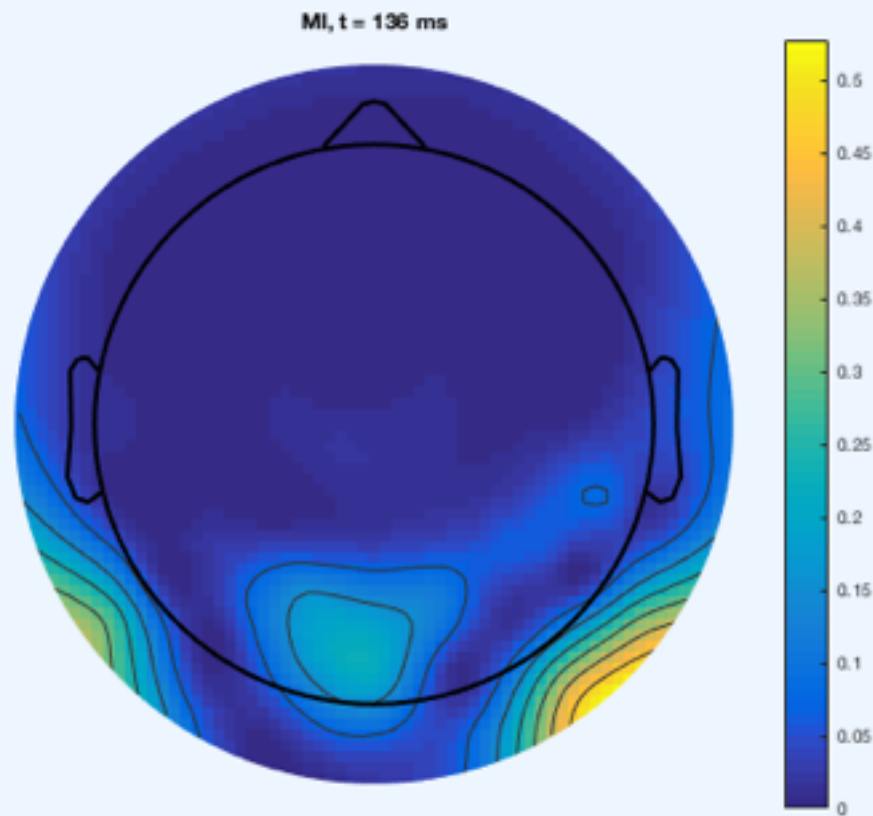
PART B:

- Split across participants.
- If your birthday is the **first half of the year**, leave `mi_model_gd` uncommented.
- If your birthday is in the **second half of the year**, comment out `mi_model_gd` and uncomment `mi_mixture_gd`

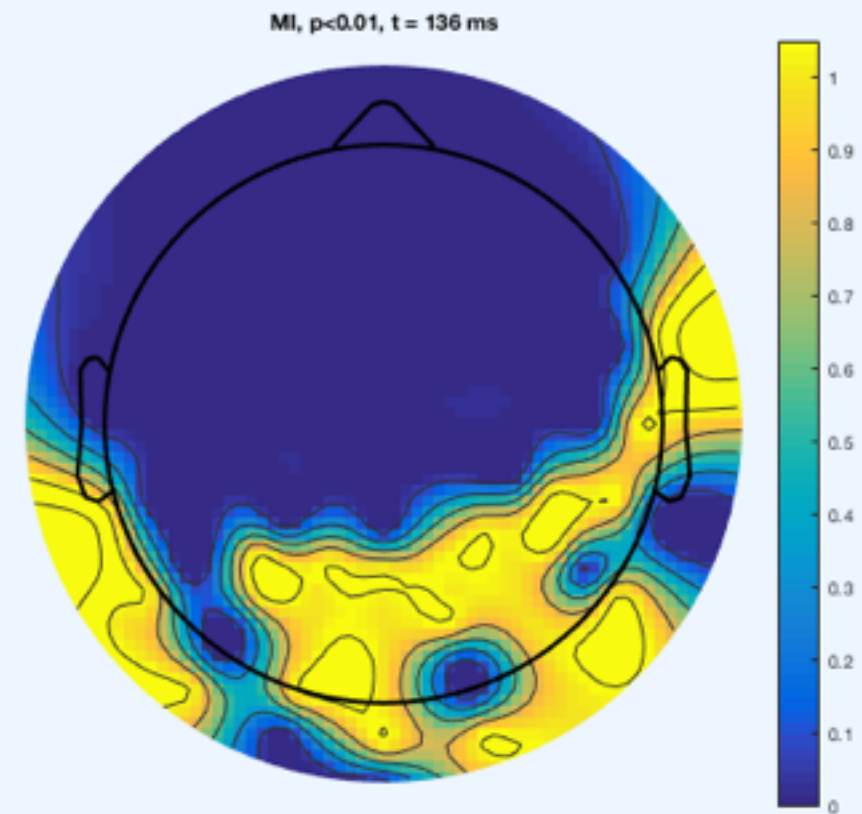
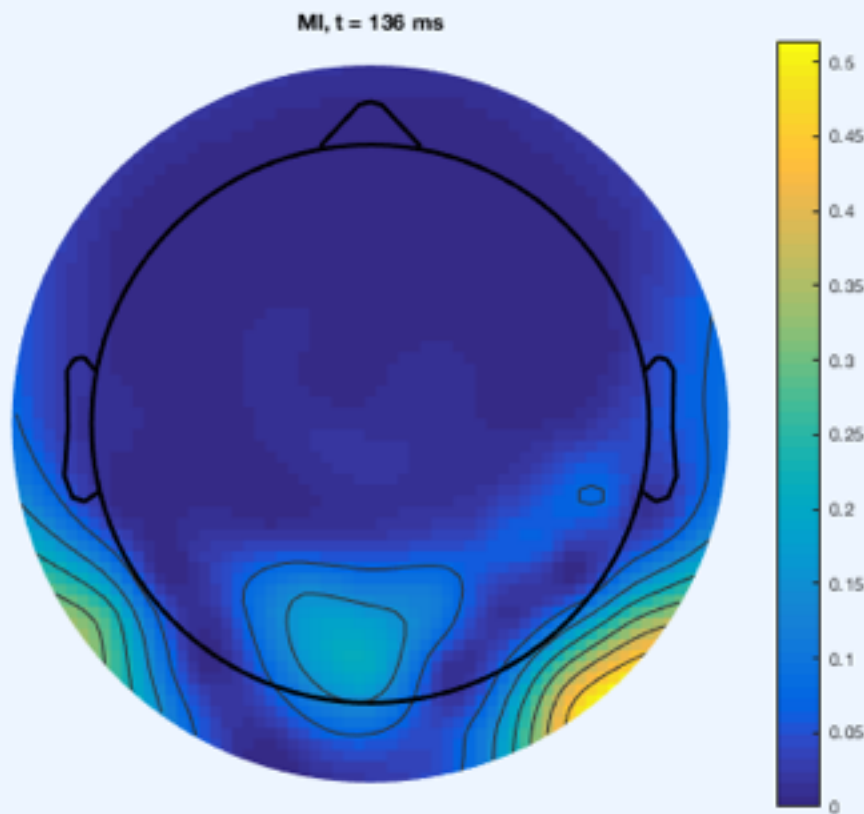
GCM I vs t-test



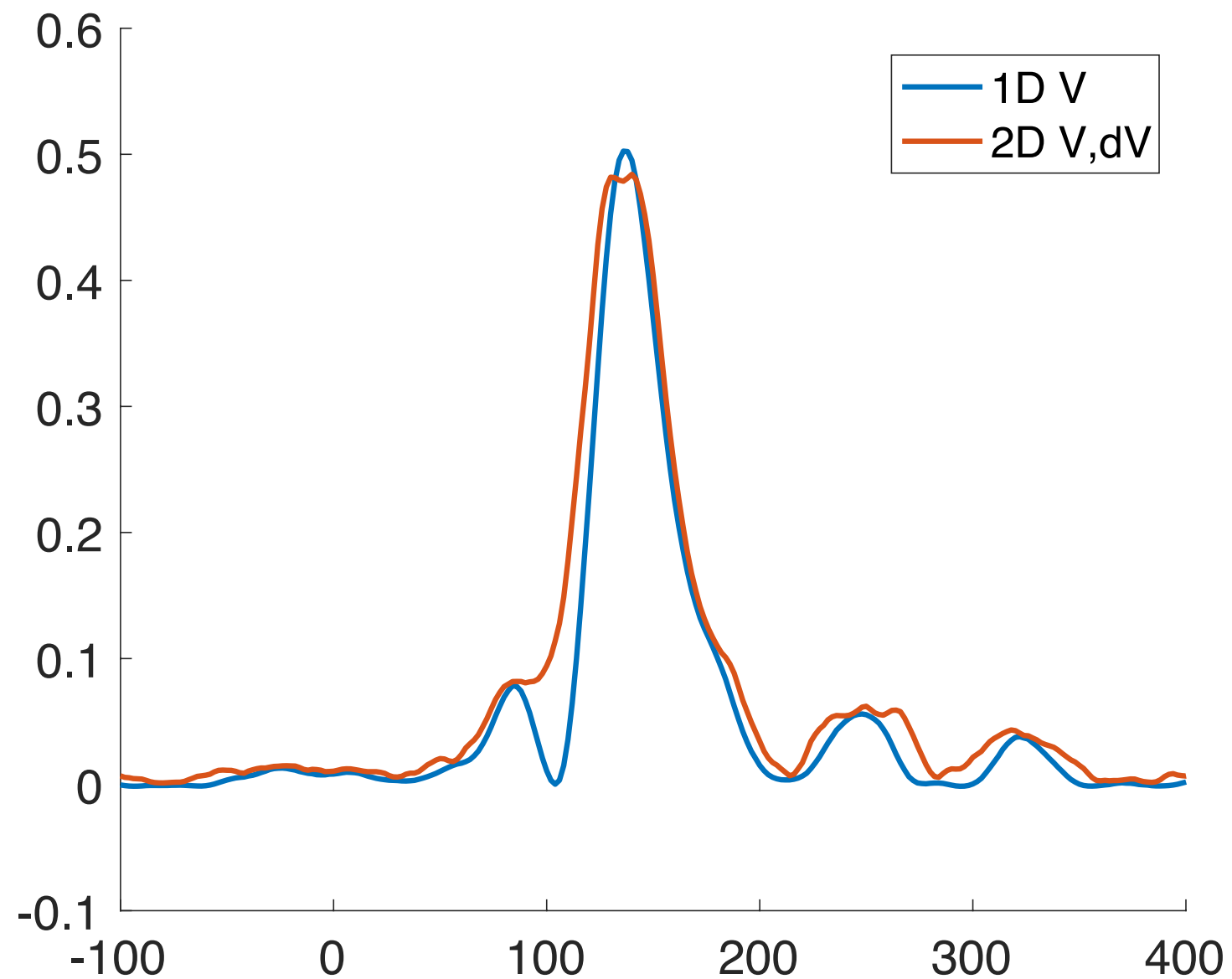
Model vs Mixture



Model vs **Mixture**



PART C: Multivariate Responses

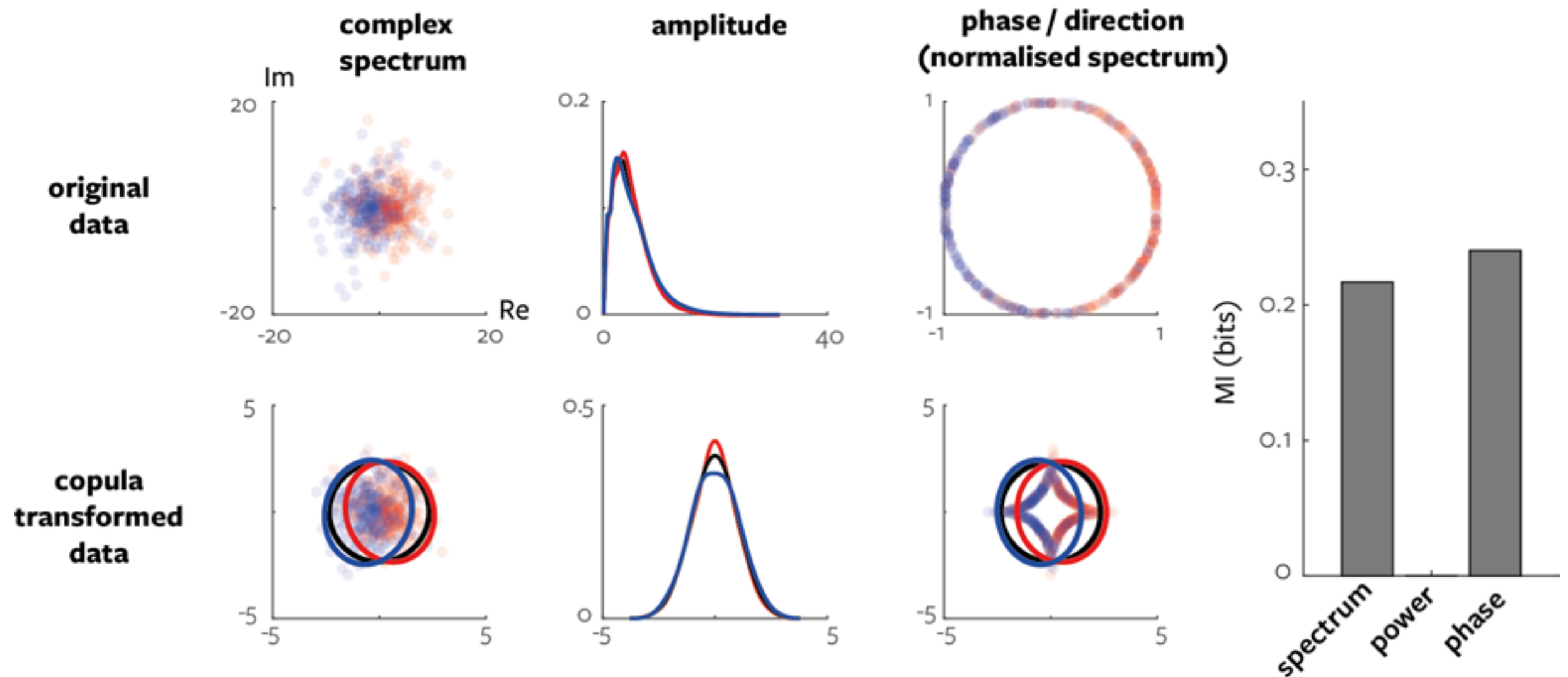


BONUS ROUND: Spectral analysis

- Stockwell transform - adaptative time-frequency representation (like wavelets)

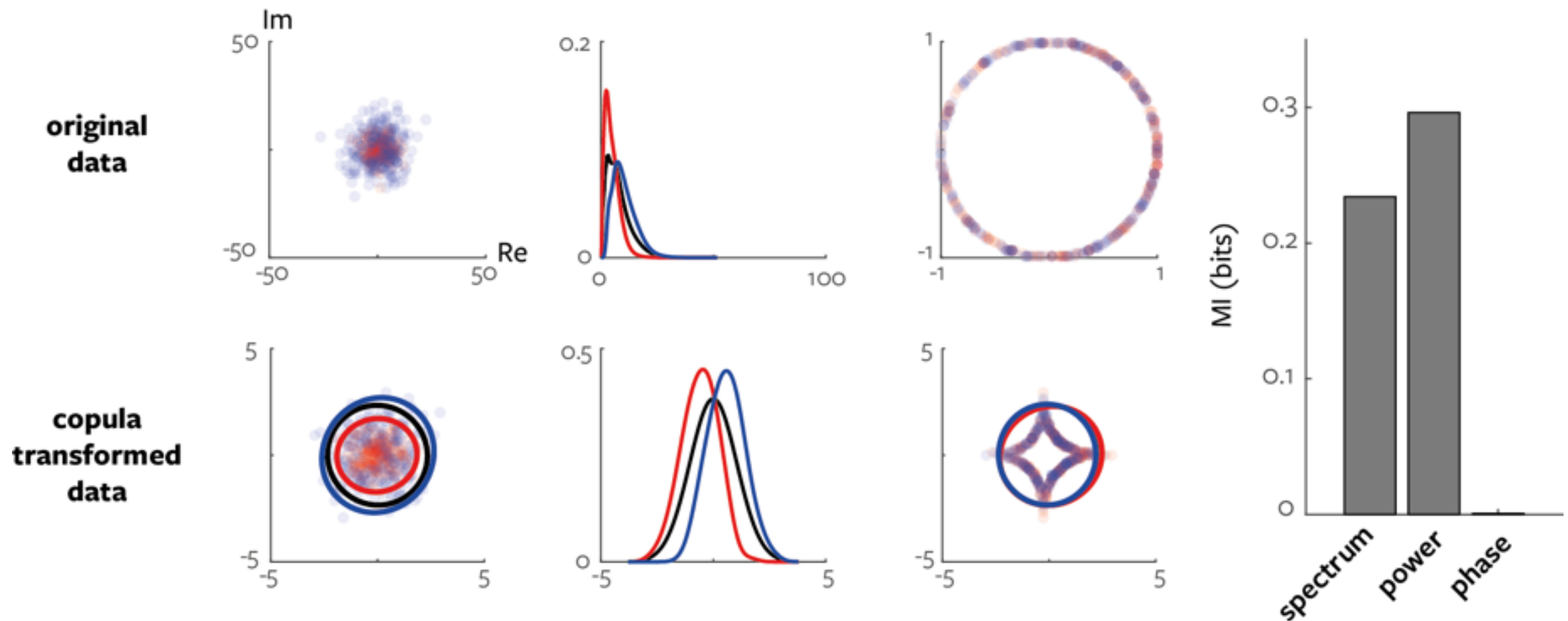
Spectral MI: phase and power

Simulation 1: Phase Modulation

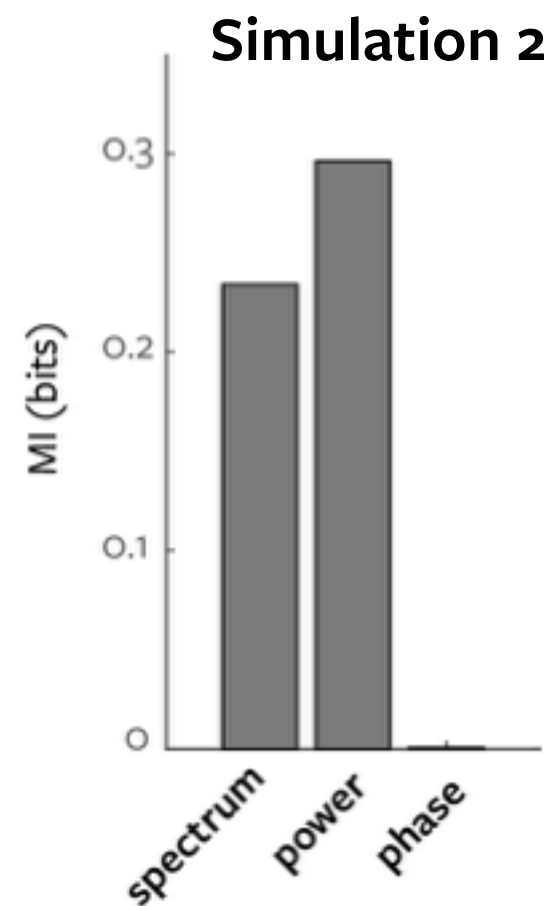
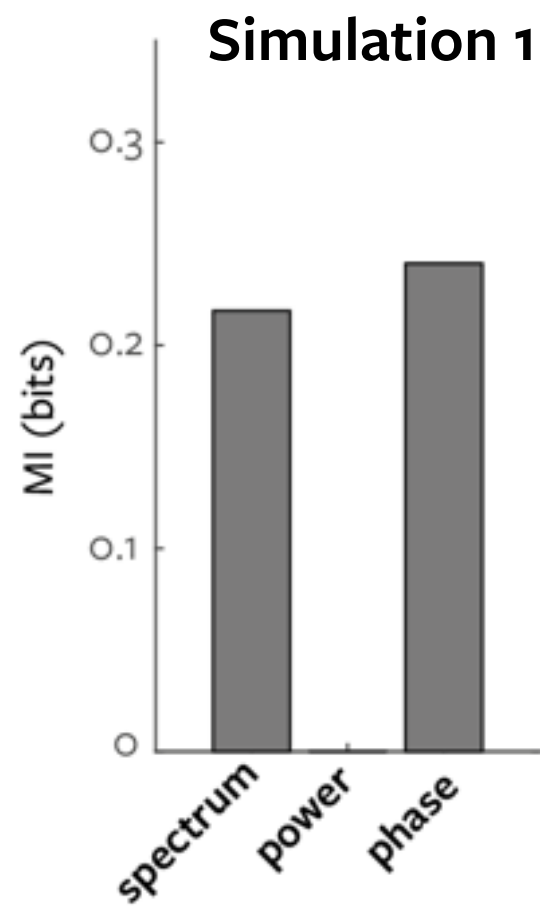
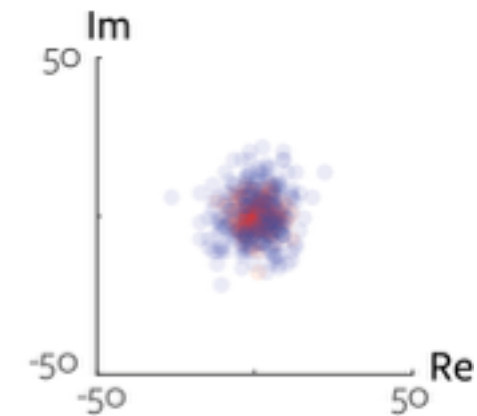
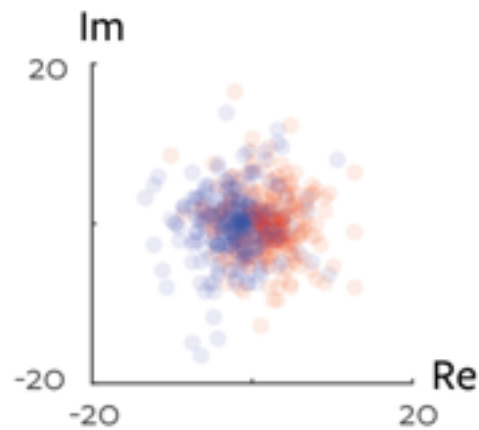


Spectral MI: phase and power

Simulation 2: Power Modulation



Spectral MI: phase and power

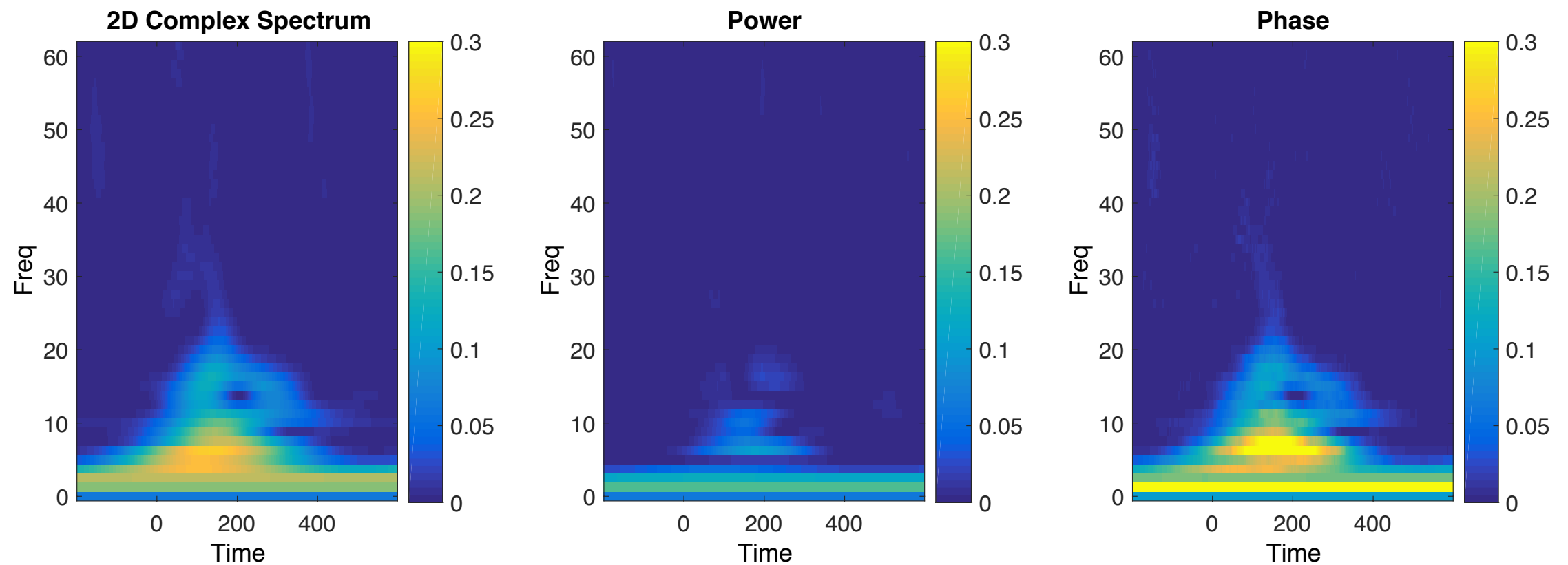


Spectral MI: phase and power

- Avoid issue of circular variables by remaining in 2D complex plane but normalising away effect of amplitude
- A test for modulation of phase + power by discrete or continuous experimental factors with a directly comparable effect size
- Can be applied to spectral data from any decomposition method (Hilbert, wavelets, empirical mode decomposition etc.)
- Interaction information : can directly relate modulations of phase and power within and across bands

GCMI Spectra

- Phase and power with directly comparable effect size



Break

Questions?

Practical 2

Continuous feature : “Bubbles” sampling

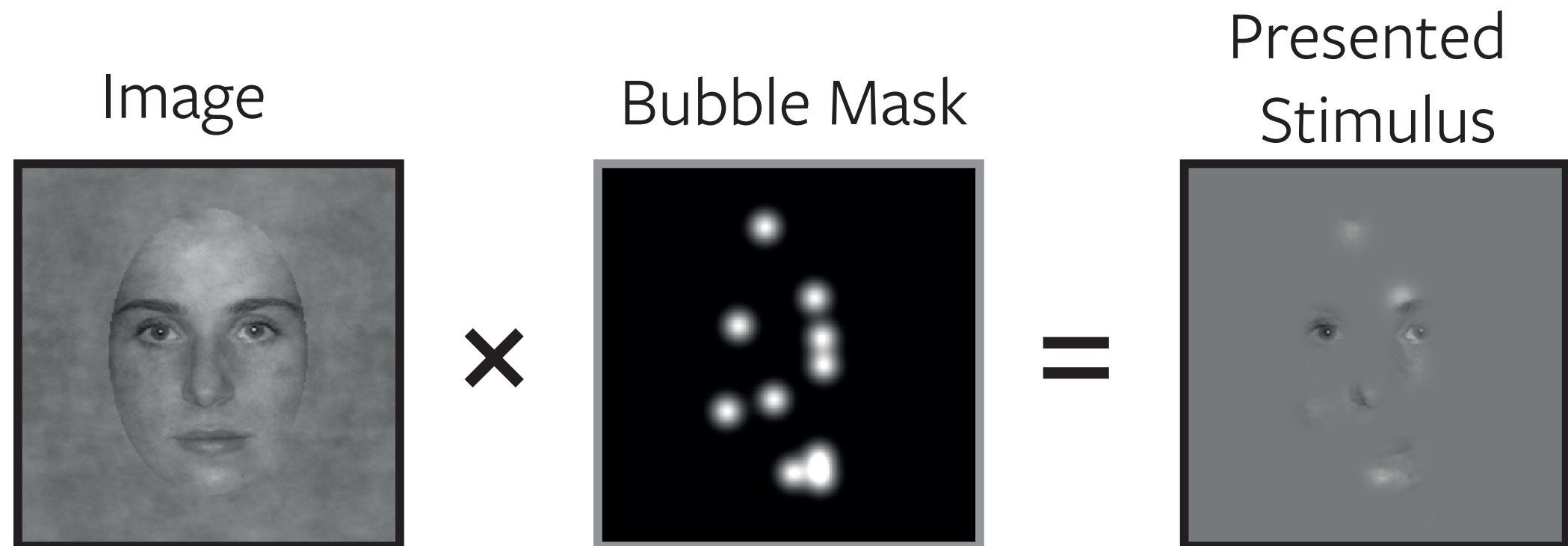
prac2_face_bubbles.m

Sampling

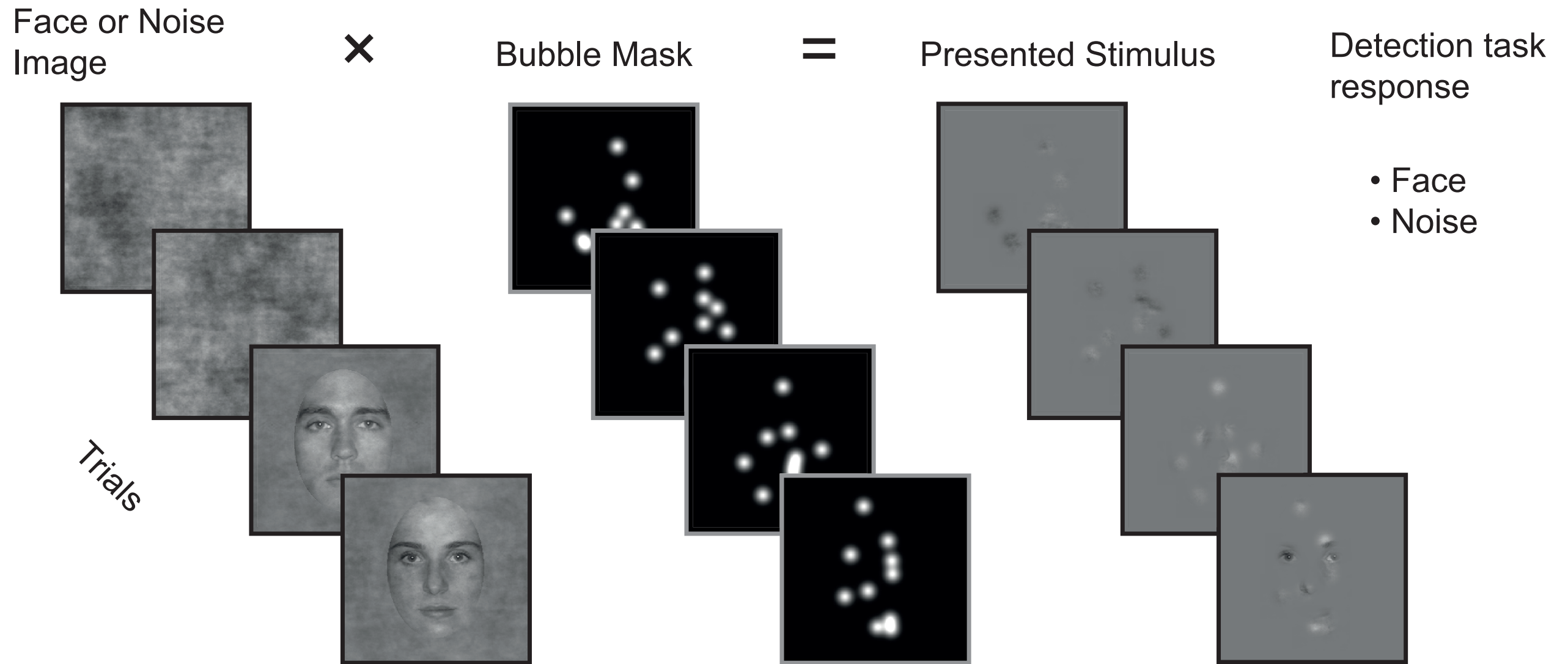
- **Random sampling** as part of experimental design
- **Generative models** can provide a tractable stimulus feature space to sample
- Variations in **dynamic naturalistic stimuli**
- Relate sampled stimulus variation to both behavioural responses and neuroimaging responses

Sampling with Bubbles

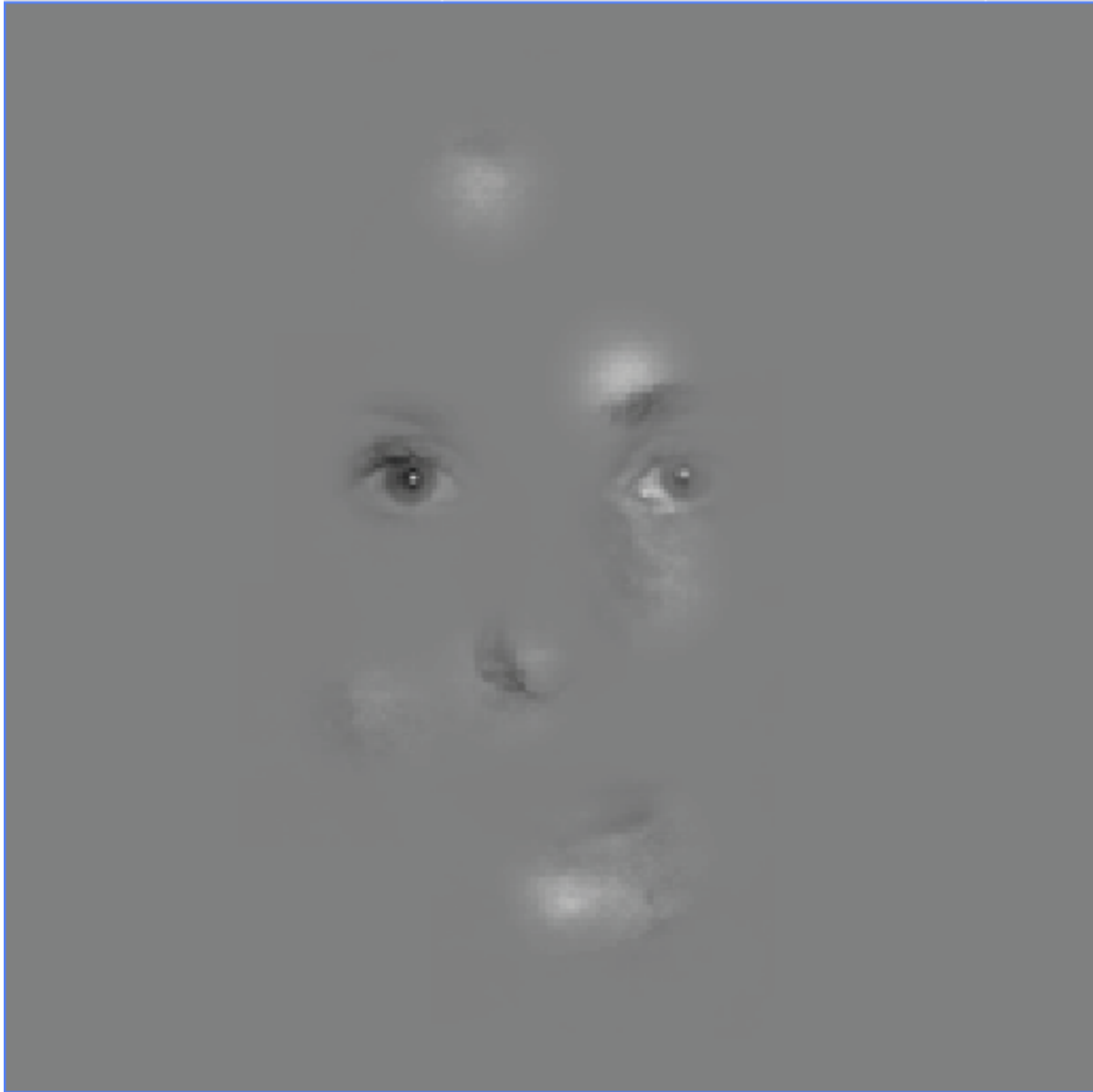
- “2d Bubbles” Direct spatial sampling of an image with randomly positioned Gaussian apertures
- Gosselin, F, and Schyns, P. “Bubbles: A Technique to Reveal the Use of Information in Recognition Tasks.” *Vision Research* 41, no. 17 (August 2001): 2261–71

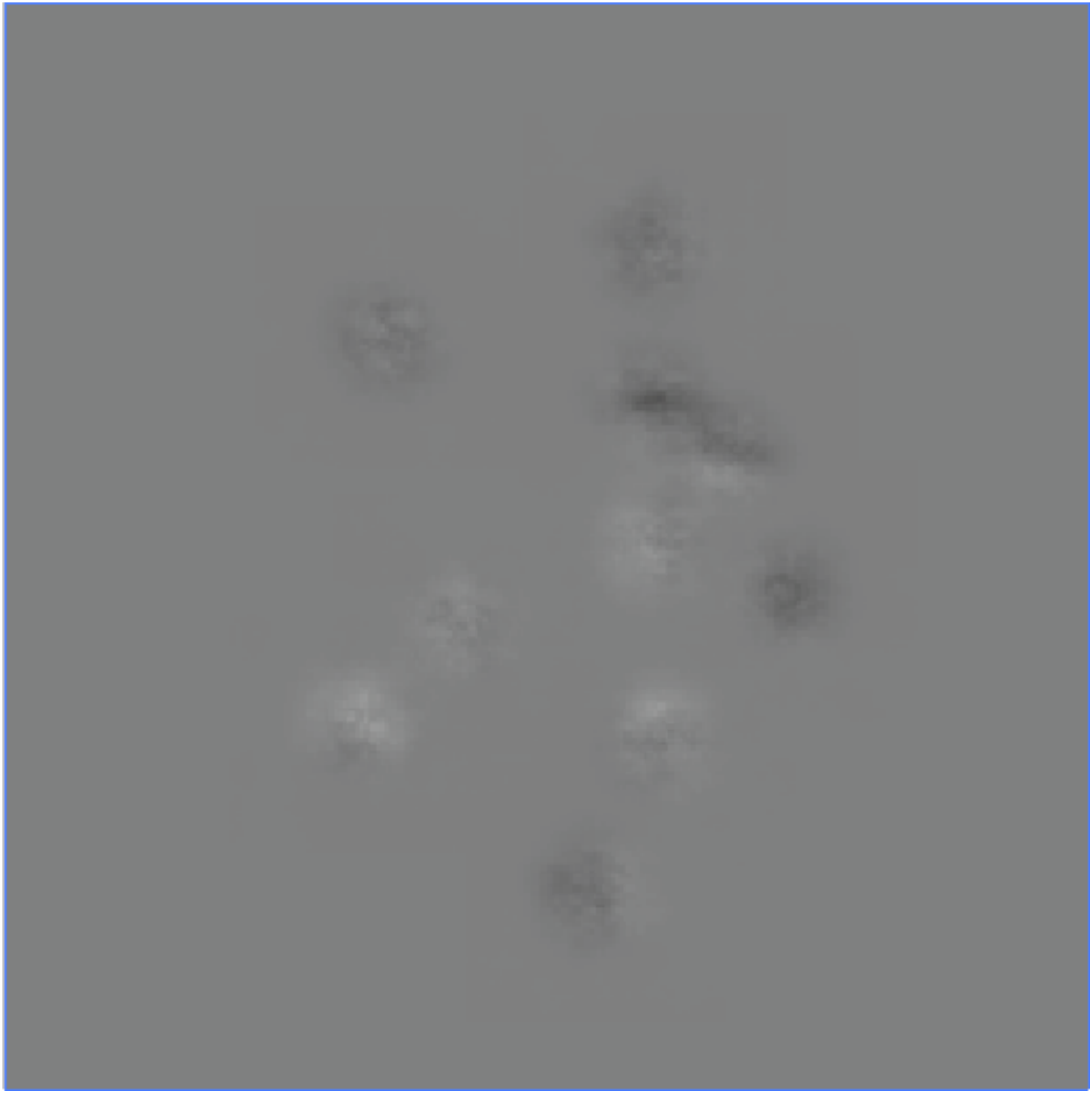


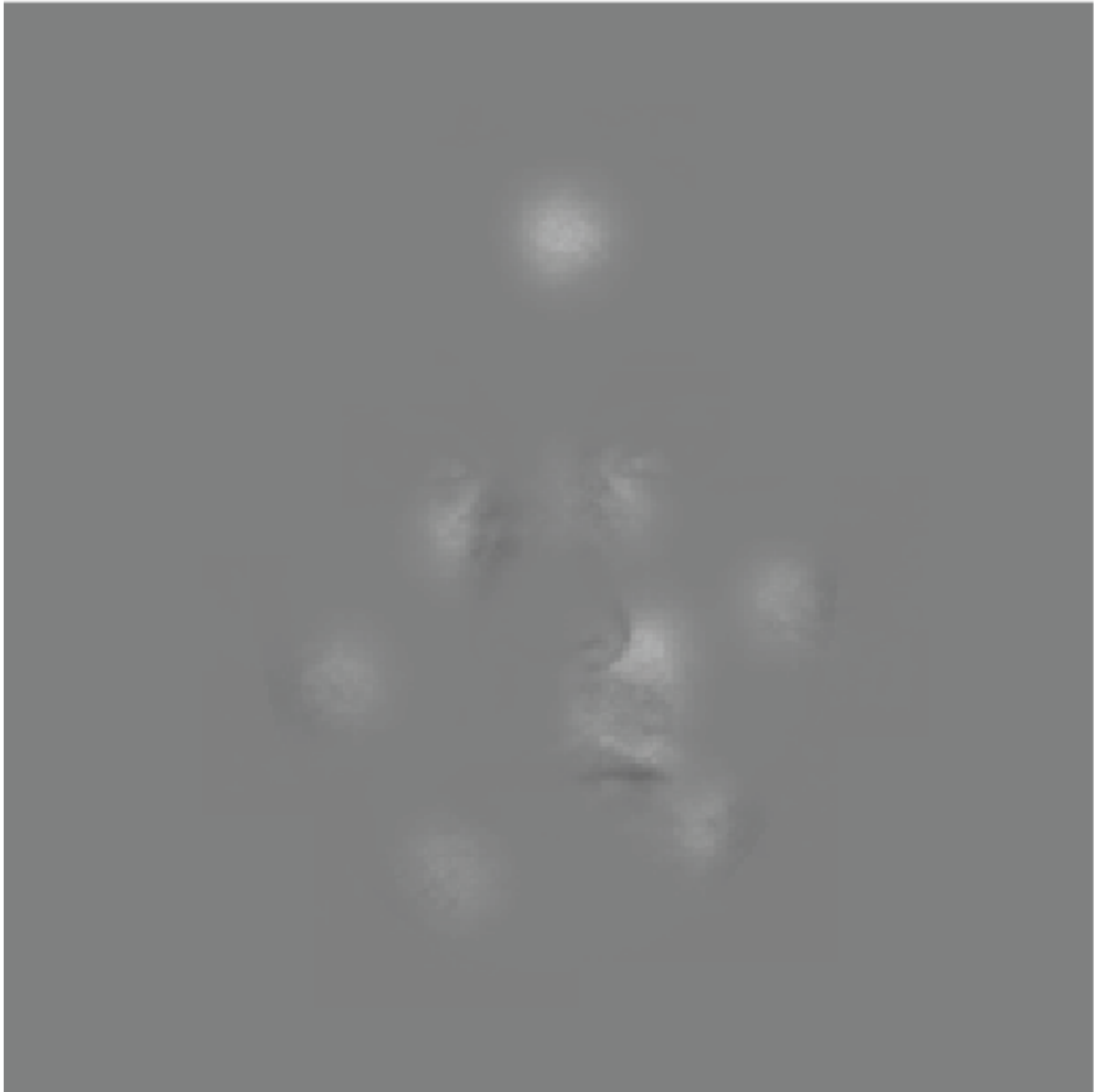
Bubbled Face Detection

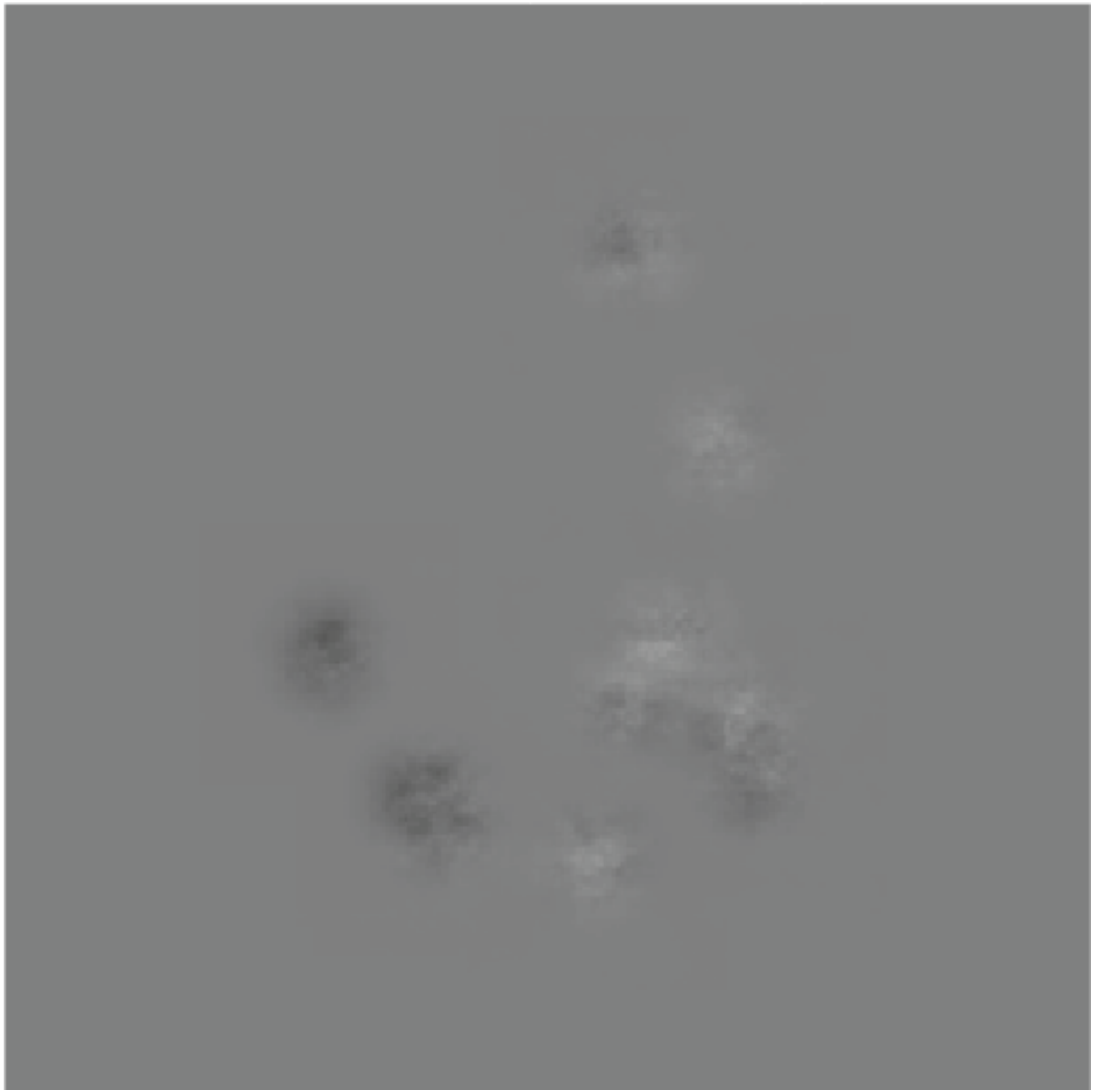


Rousselet, G. A., Ince, R. A. A., van Rijsbergen, N. J., and Schyns, P. G. (2014). Eye coding mechanisms in early human face event-related potentials. *J Vis* 14, 7.







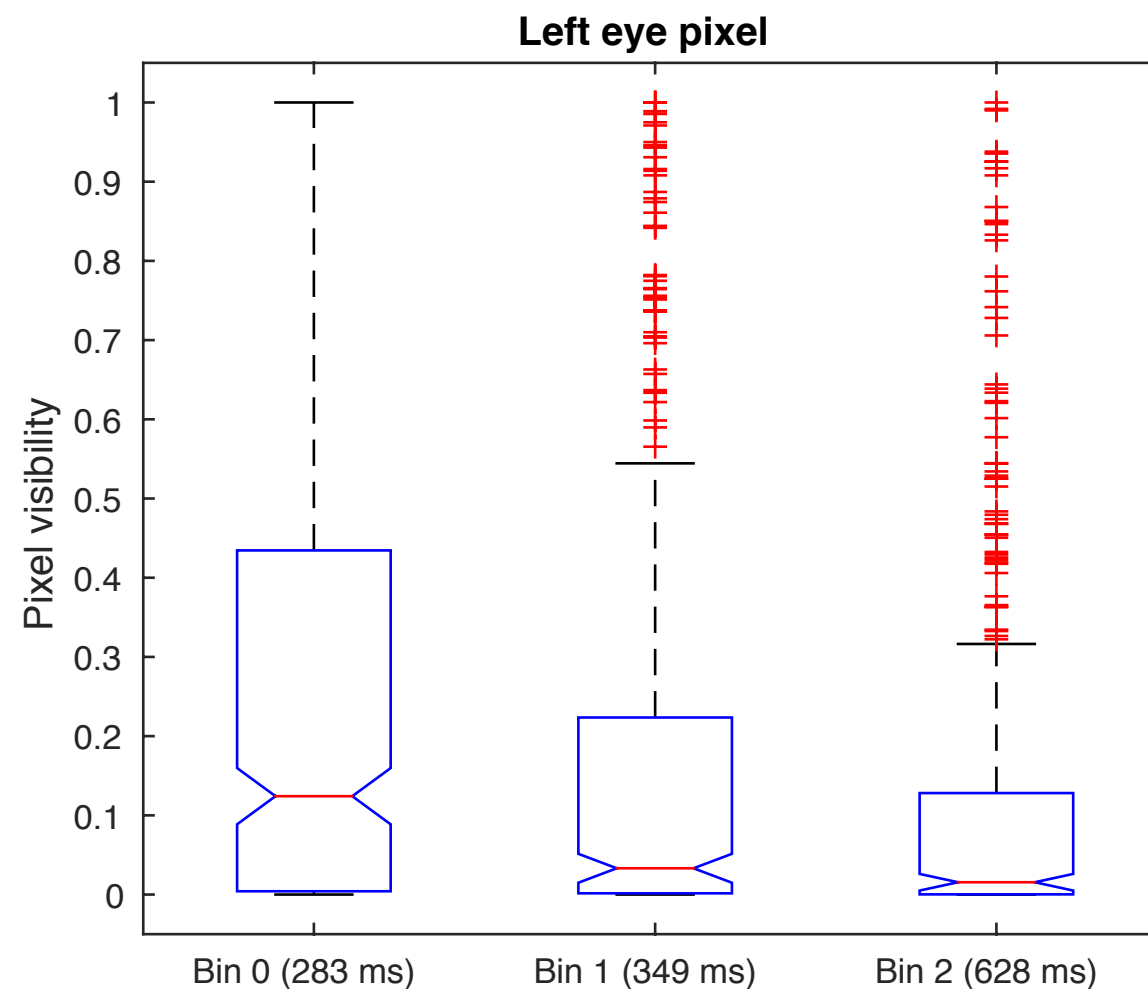


Data

- **Stimuli:** Bubble mask images
[trials x vertical pixels x horizontal pixels]
- **Behavioural Responses:** face vs noise, reaction time
[trials]
- **EEG:** CSD + bandpass filter
[trials x time points x sensors]
- **Challenge:** relate high-dimensional stimulus to behaviour and high-dimensional EEG response

Part A: Pixels vs Reaction Time

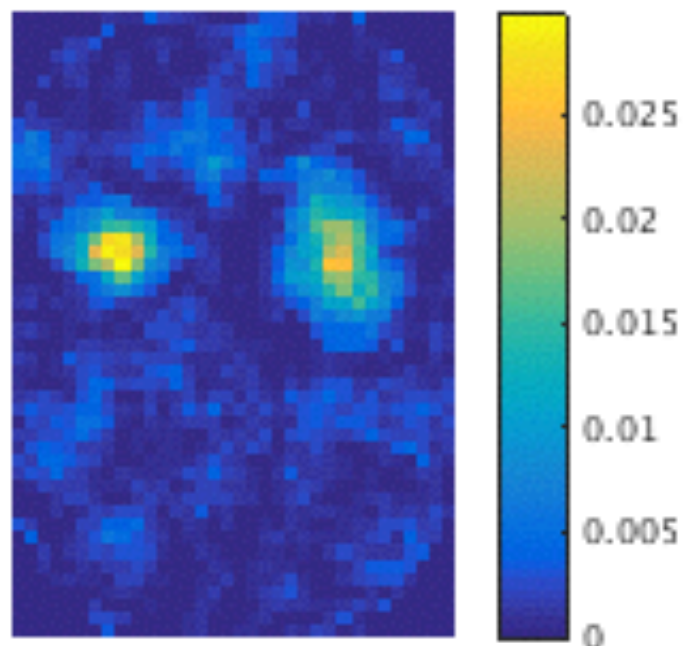
- **Reverse correlation:** correlate each pixel's visibility with a behavioural response



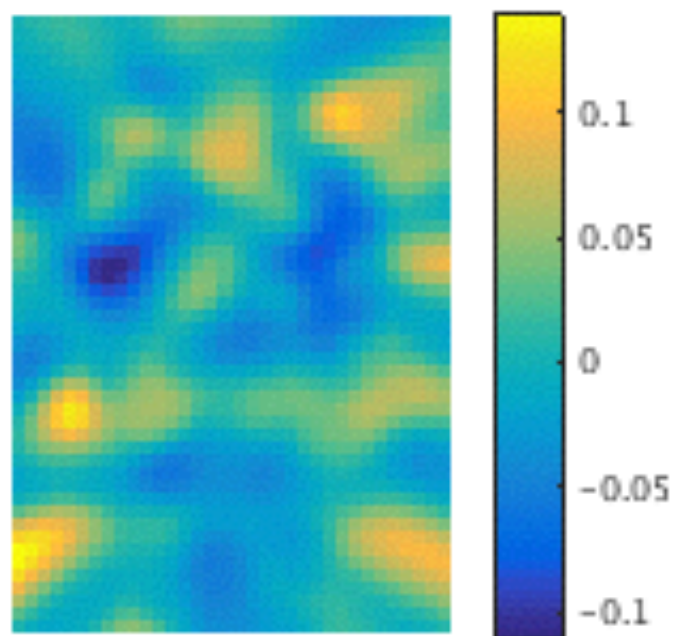
Part A: Pixels vs Reaction Time

- Calculate MI independently for each pixel; plot resulting image
- Scale gives good contrast for exploratory analysis

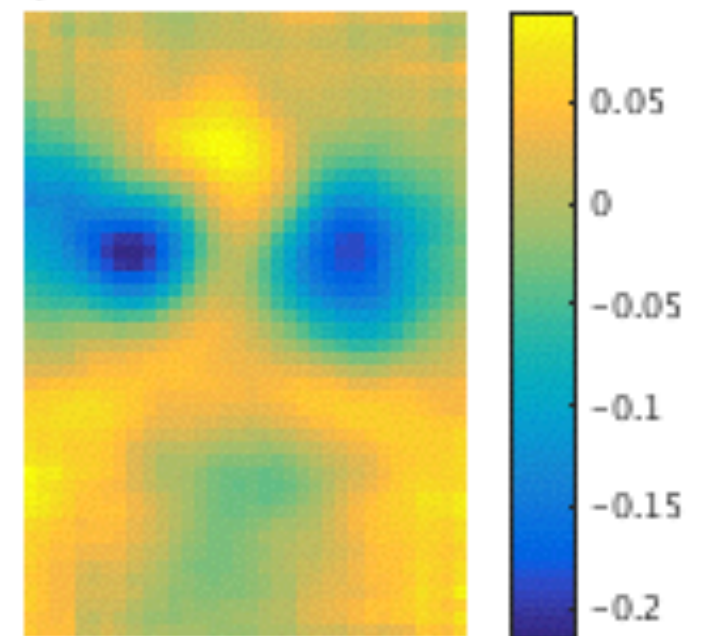
Mutual Information



Pearson Correlation



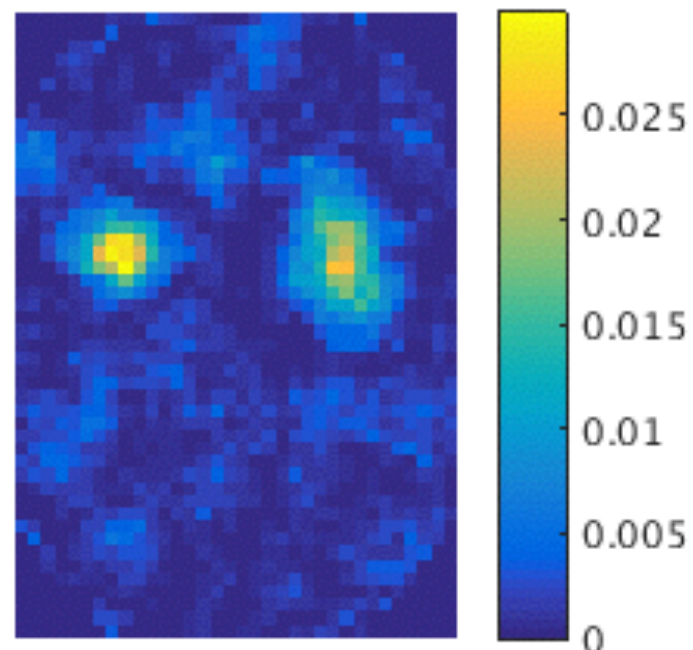
Spearman Correlation



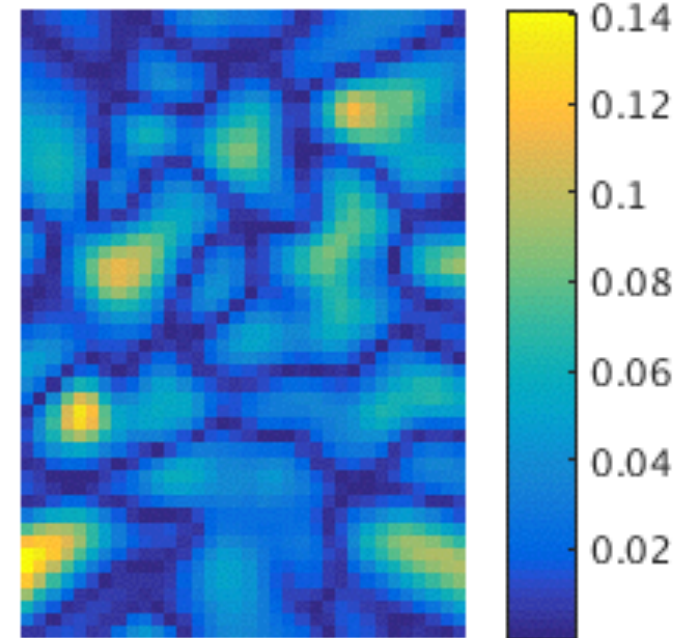
Part A: Pixels vs Reaction Time

- Calculate MI independently for each pixel; plot resulting image
- Scale gives good contrast for exploratory analysis

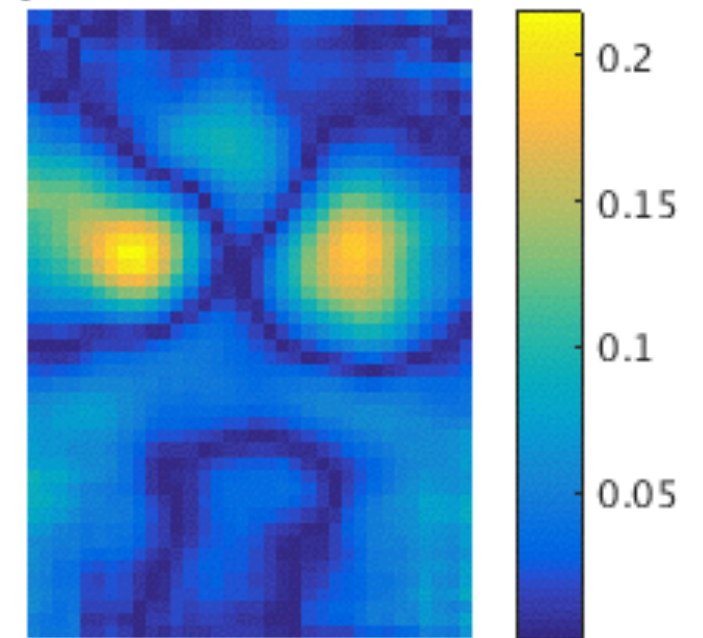
Mutual Information



Pearson Correlation

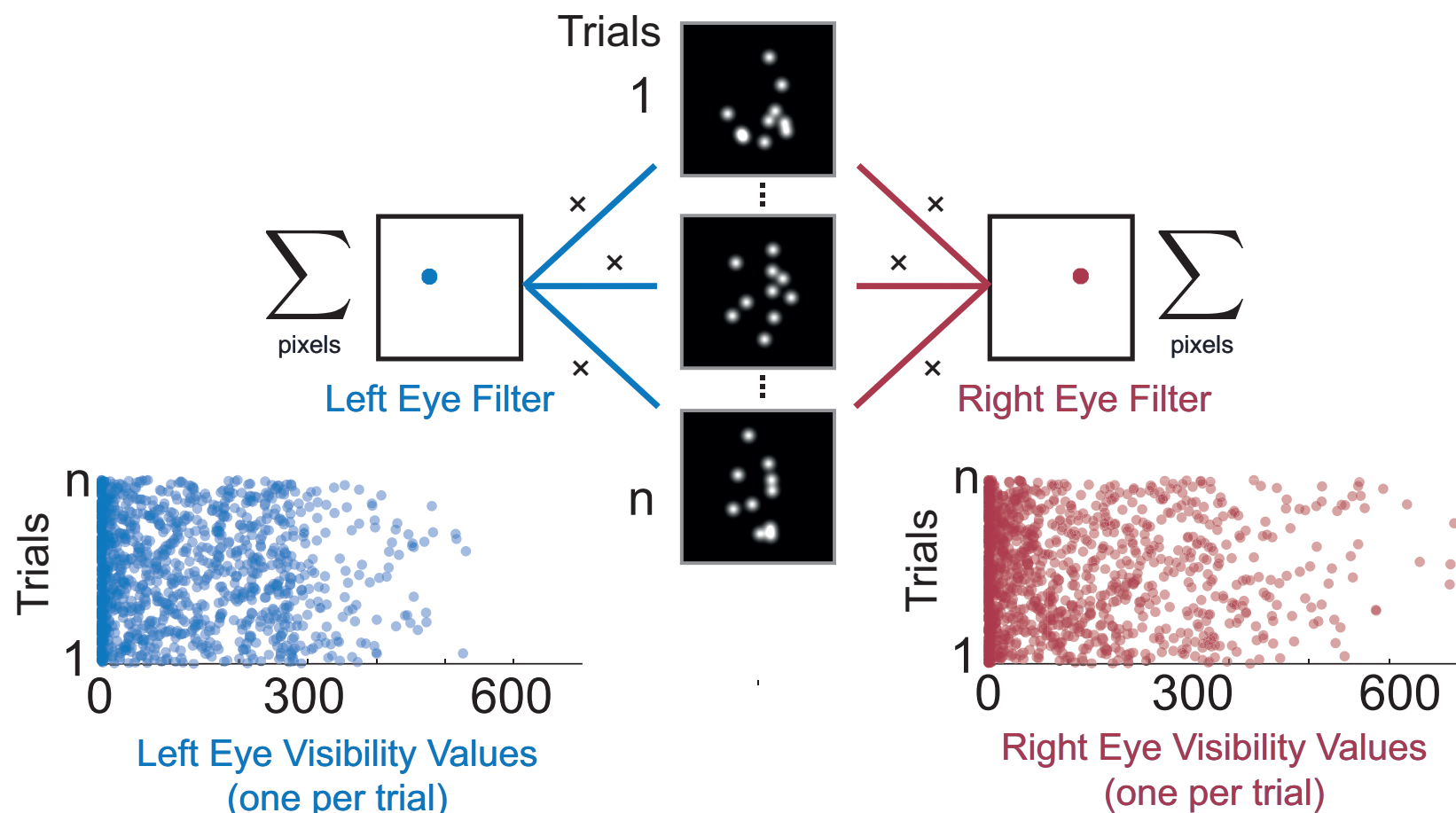


Spearman Correlation



Part B: Dimensionality Reduction

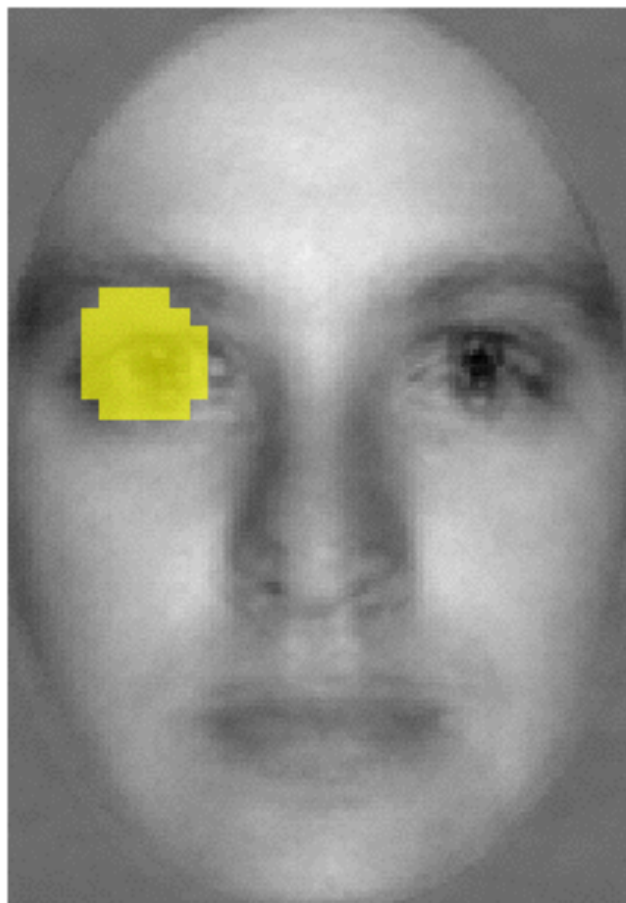
- Full pixel MI images at every sensor + time point. Computationally intensive and hard to visualise.
- Reduce dimensionality by considering the visibility of small regions rather than individual pixels
- Defined a priori, from behaviour or from data-driven methods



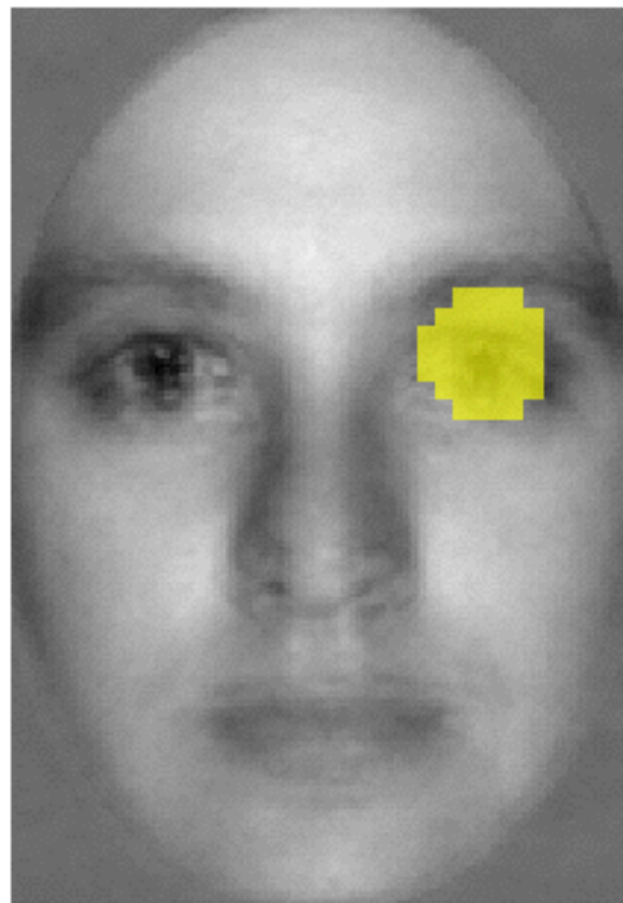
Part B: Dimensionality Reduction

- Sum bubble mask value (pixel visibility) within masked regions

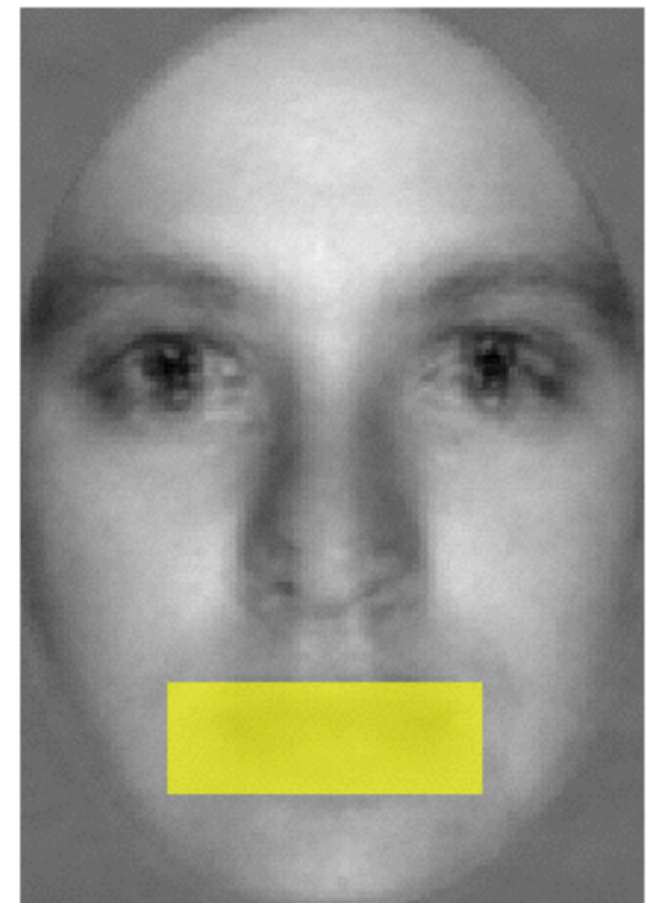
Left eye



Right eye

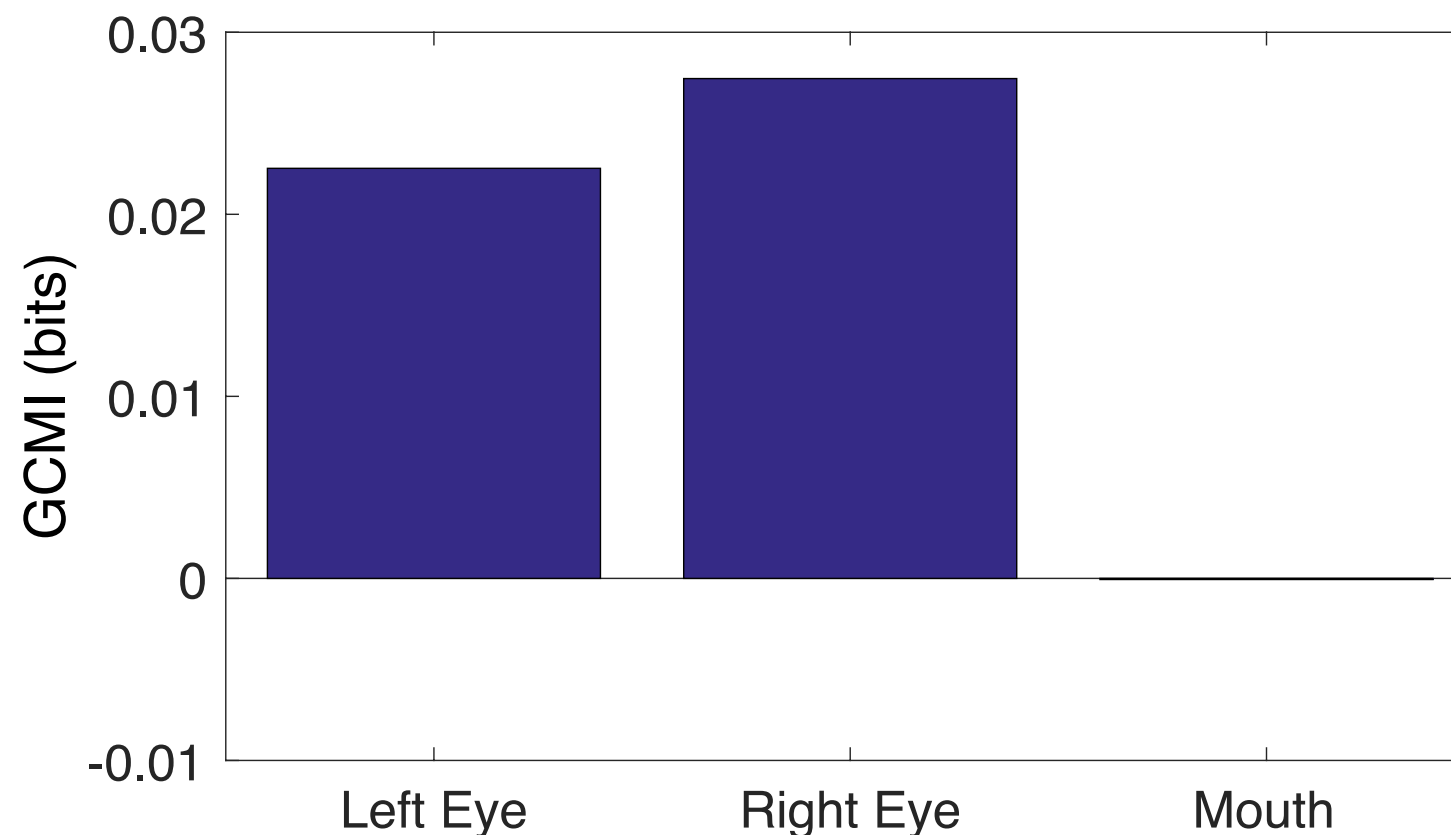


Mouth



Ex 1: Stimulus Dimensionality Reduction

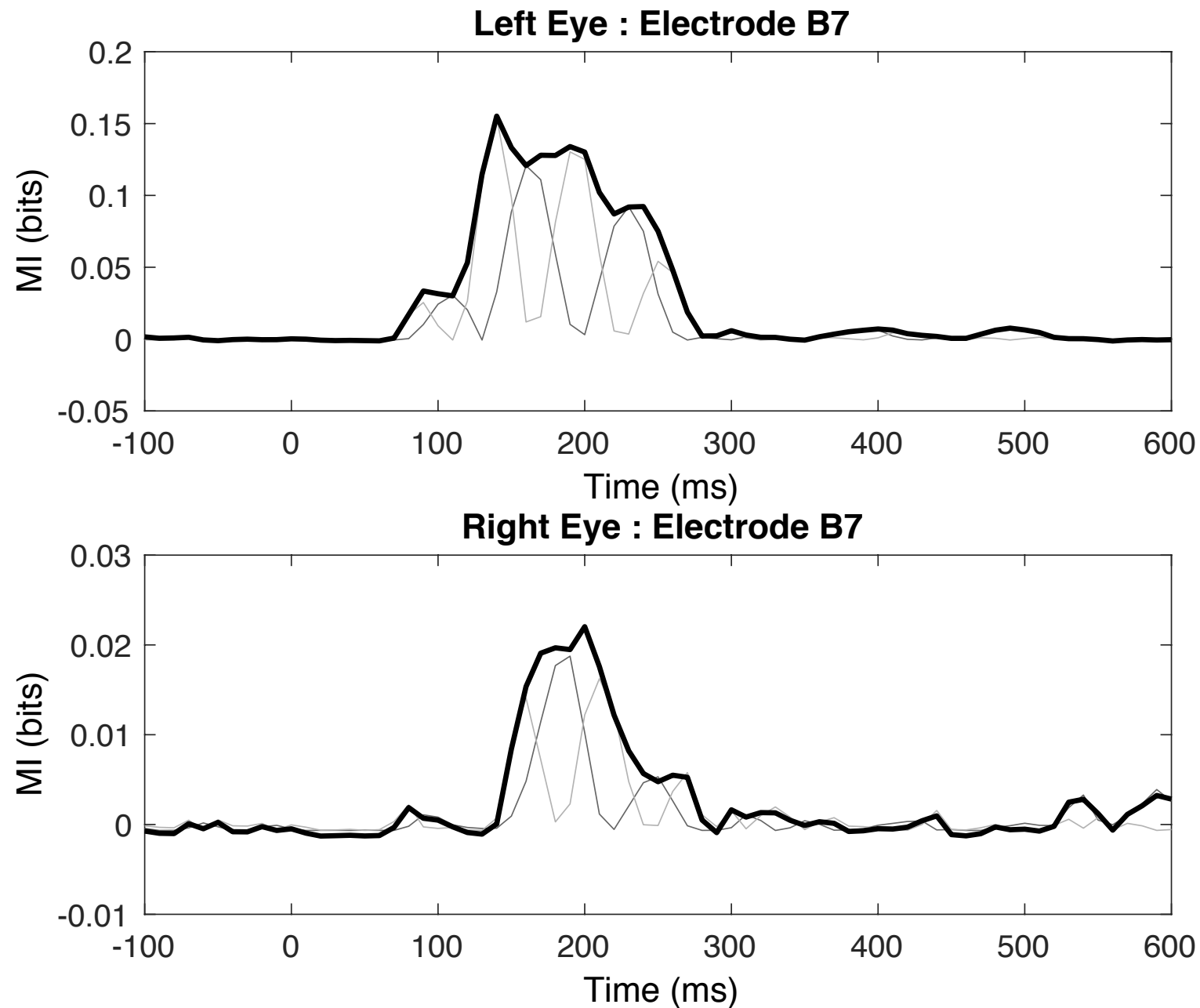
- Relationship with RT preserved in 1D stimulus feature
- Both eyes affect RT; mouth region doesn't (consistent with full MI pixel image)



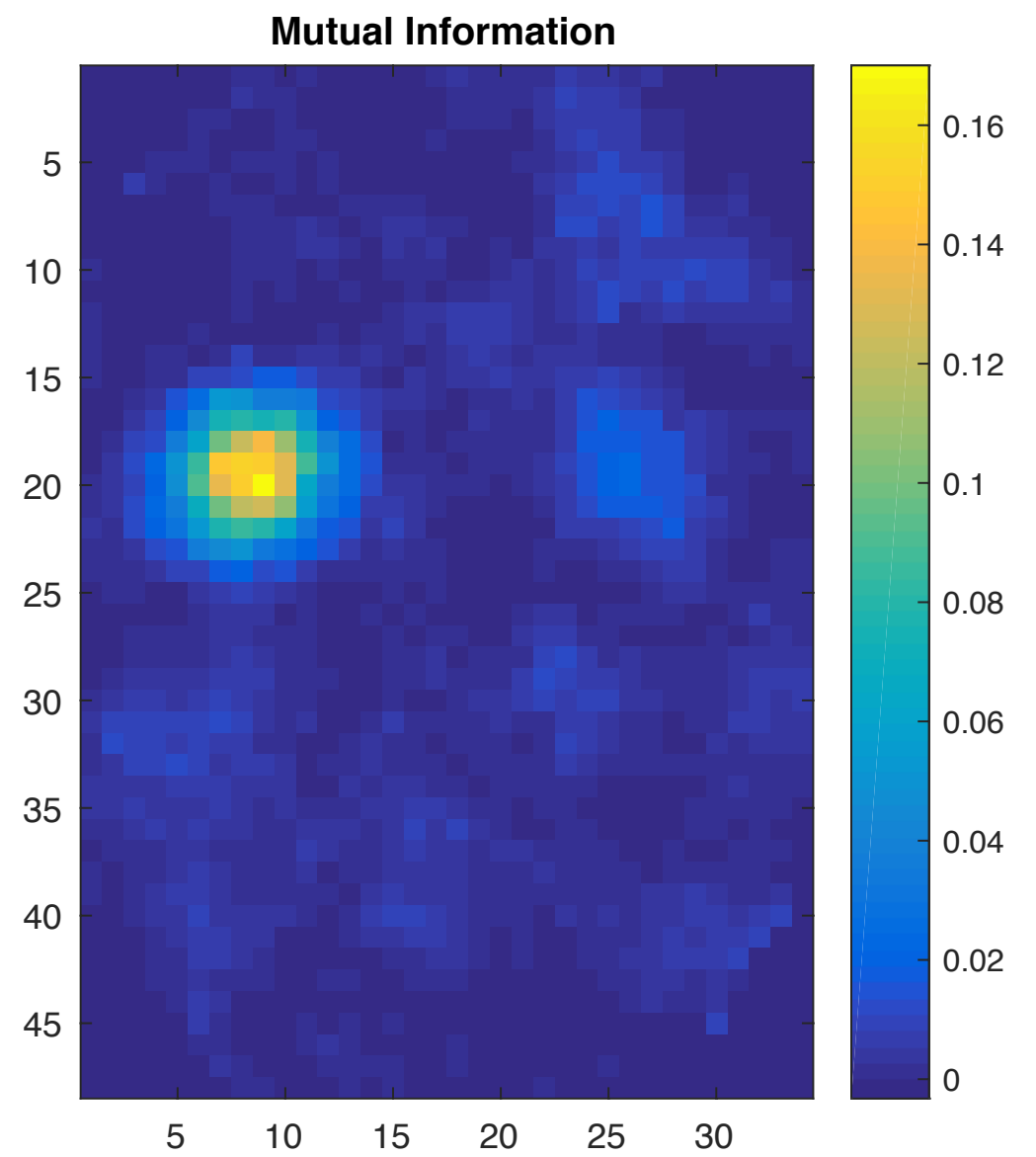
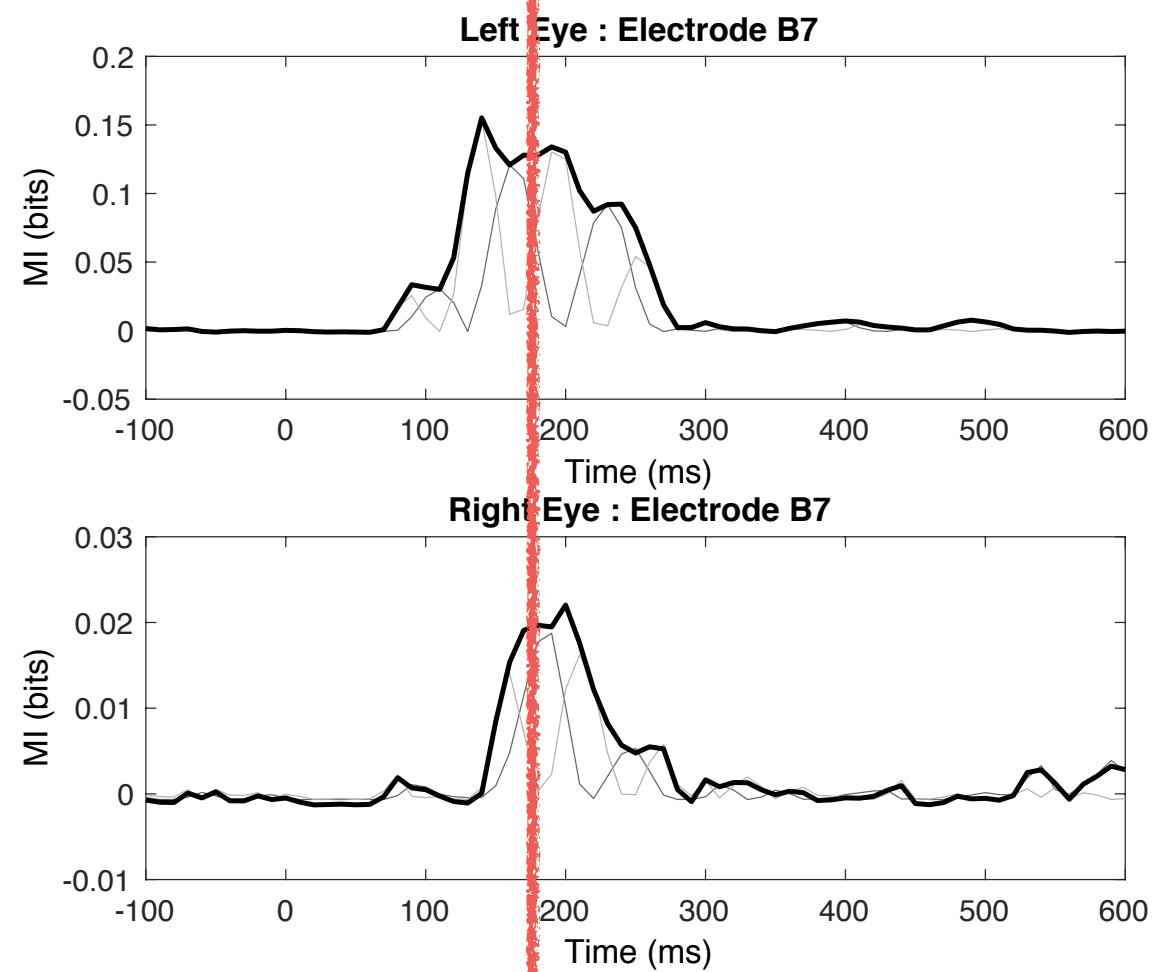
Behaviour summary

- Can do this sort of analysis with any behavioural measure: response, accuracy (correct/incorrect), choice confidence
- Can use other sampling mechanisms (generative models, noise sampling, continuous stimuli). Key requirement is diverse sampling of high-dimensional naturalistic stimulus space

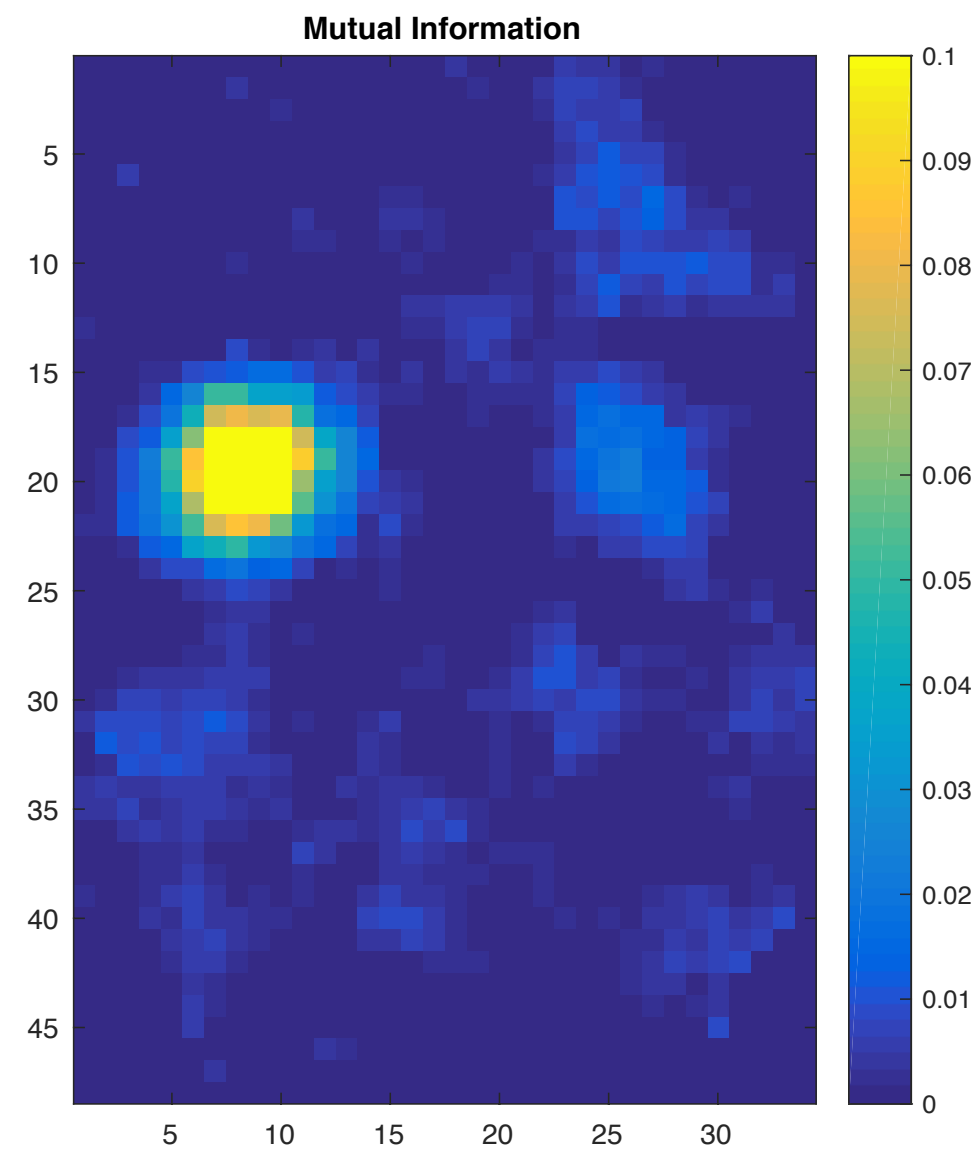
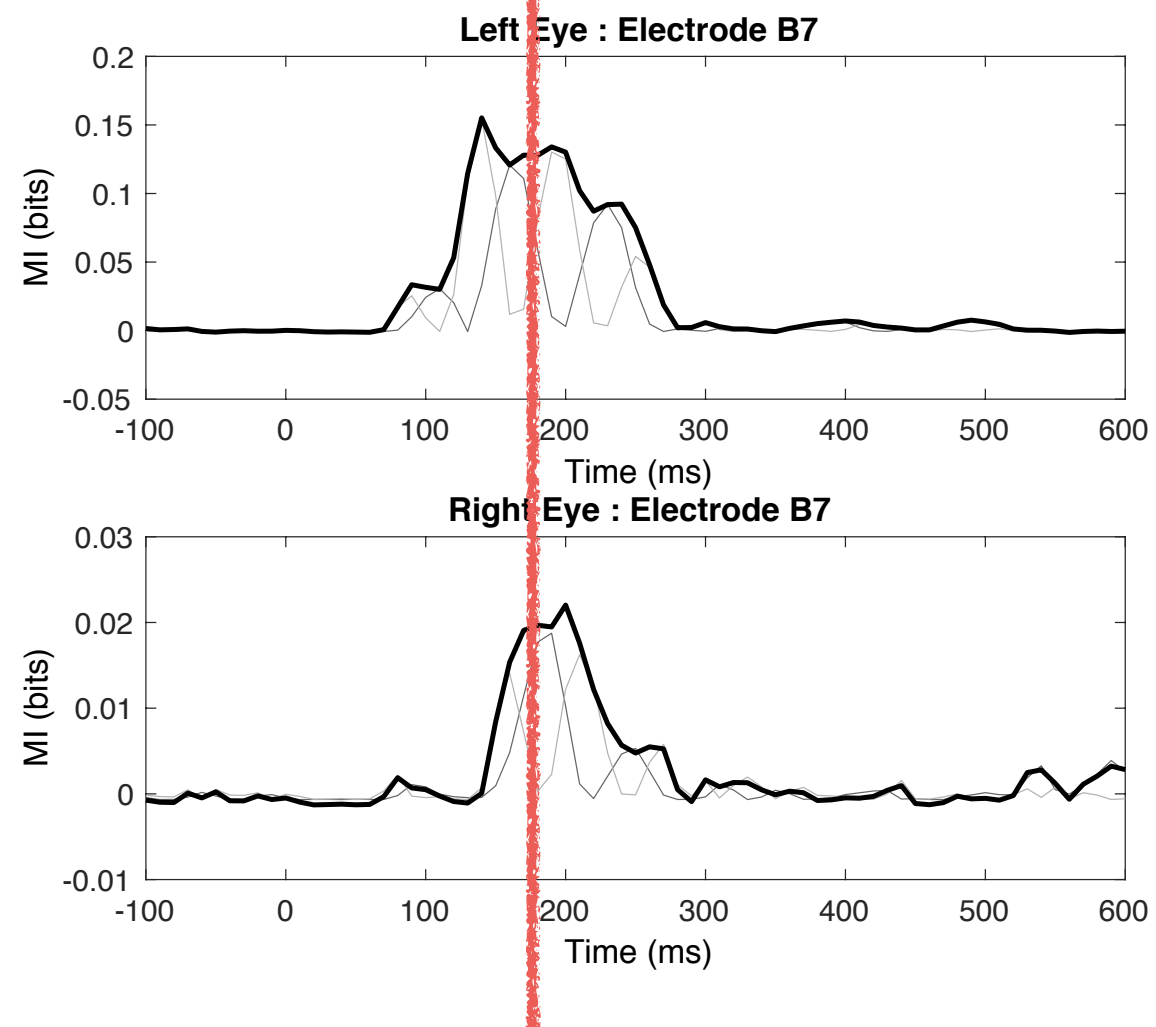
Part C: Stimulus Feature vs EEG



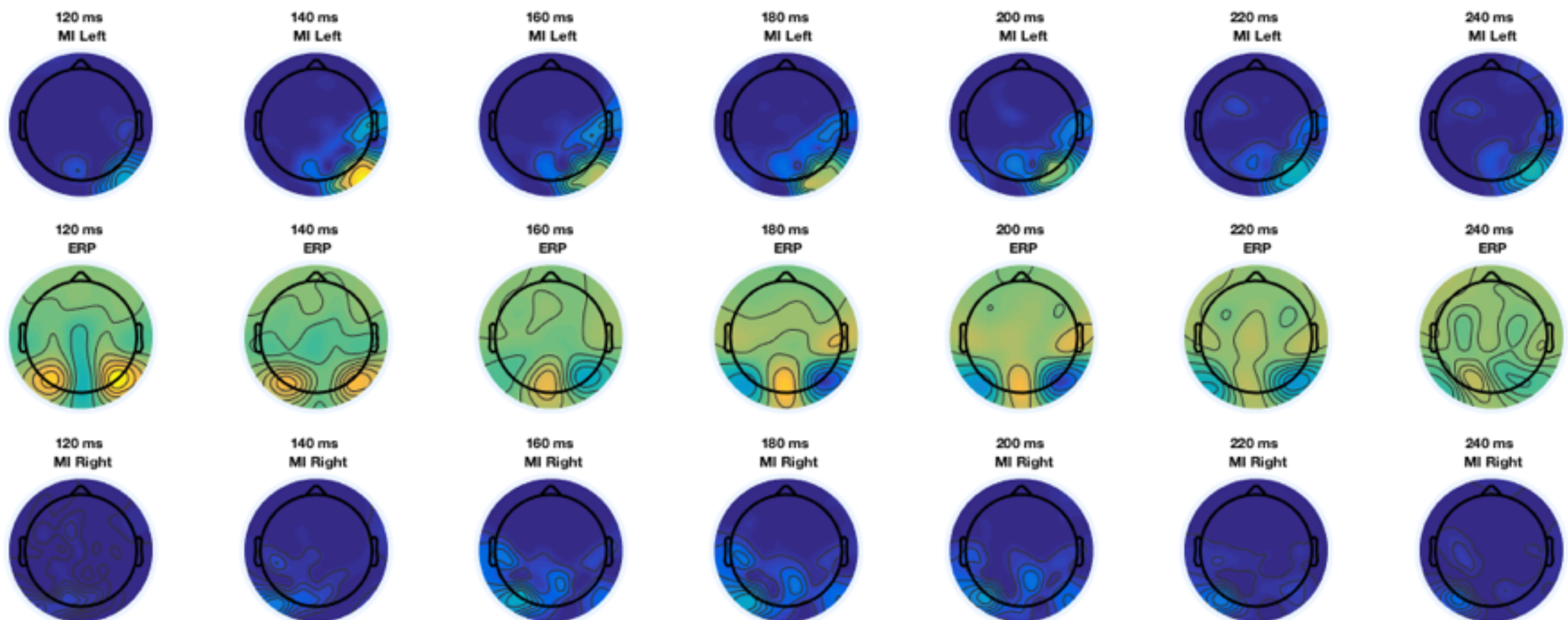
$t = 170\text{ms}$



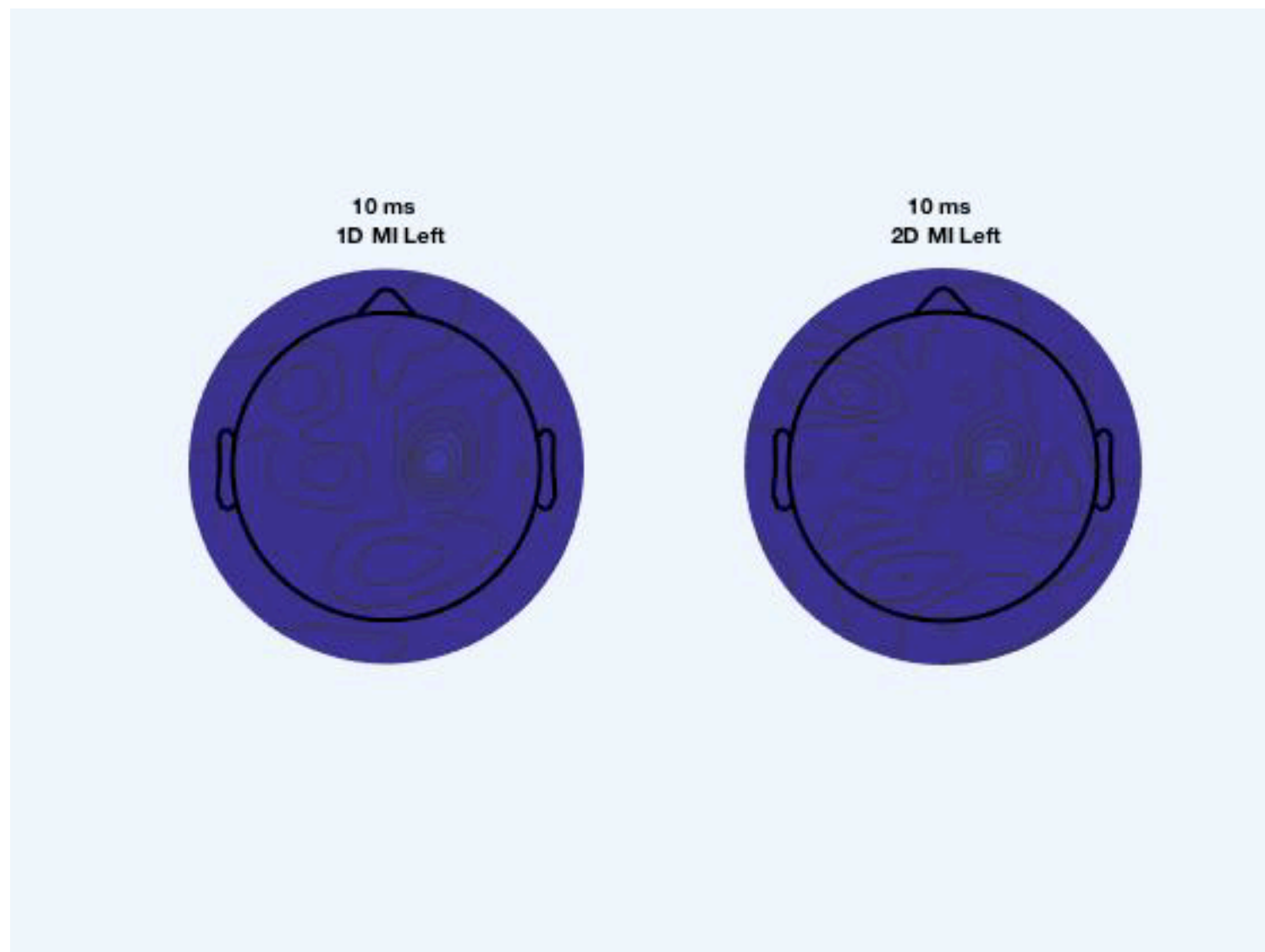
$t = 170\text{ms}$



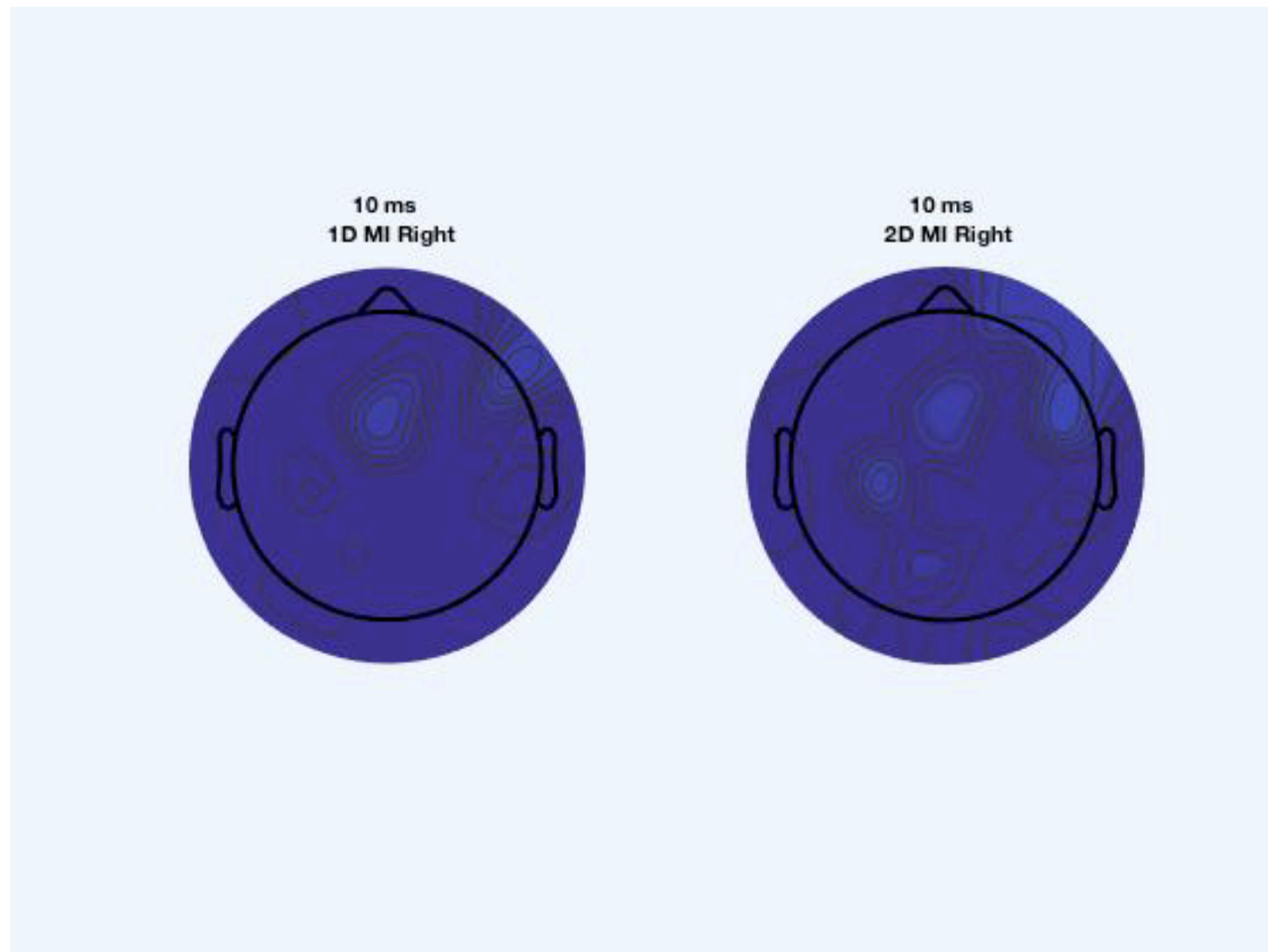
- **Activation** is symmetric (bilateral N170 ERP)
- But (stimulus) **information** is asymmetric (lateralized)
- Information (Representation, Coding, Stimulus) vs Activation



BONUS ROUND: Movies



BONUS ROUND: Movies



Recap

- Rank based statistic
- Continuous and discrete
- Multivariate (spectra, temporal derivative)

Break?

- Move to `prac3_eeg_temporal_interaction.m`

Representational Interactions

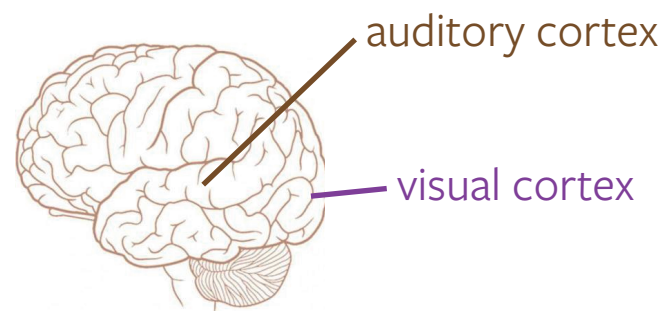
- Information processing perspective requires not just tracking stimulus modulations, but relating the representations in different neural signals
- Currently the only method that addresses this in Representational Similarity Analysis
- We can do this with information theory

Representational Interactions

Neural Responses

Spatial Regions

beamformed MEG activity in:



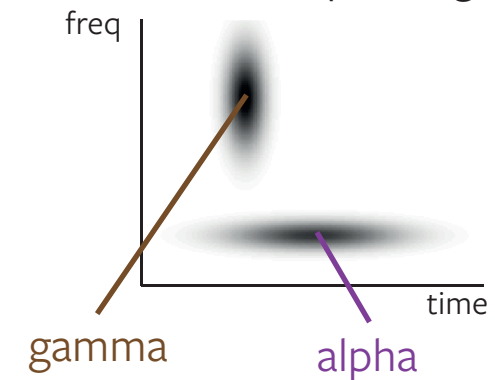
Temporal Regions

stimulus modulation of evoked signal on parietal EEG electrode

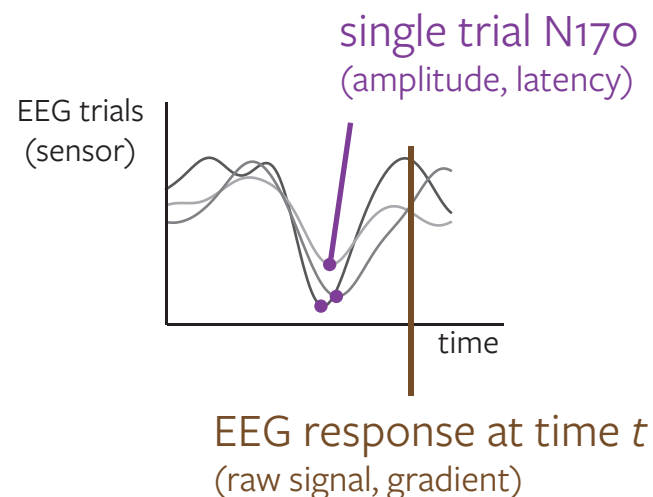


Frequency Regions

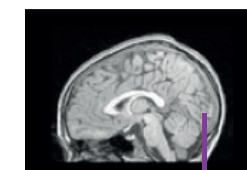
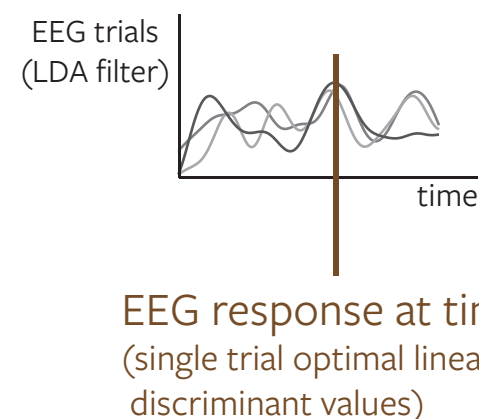
MI of MEG spectrogram



Reduced Response Descriptions

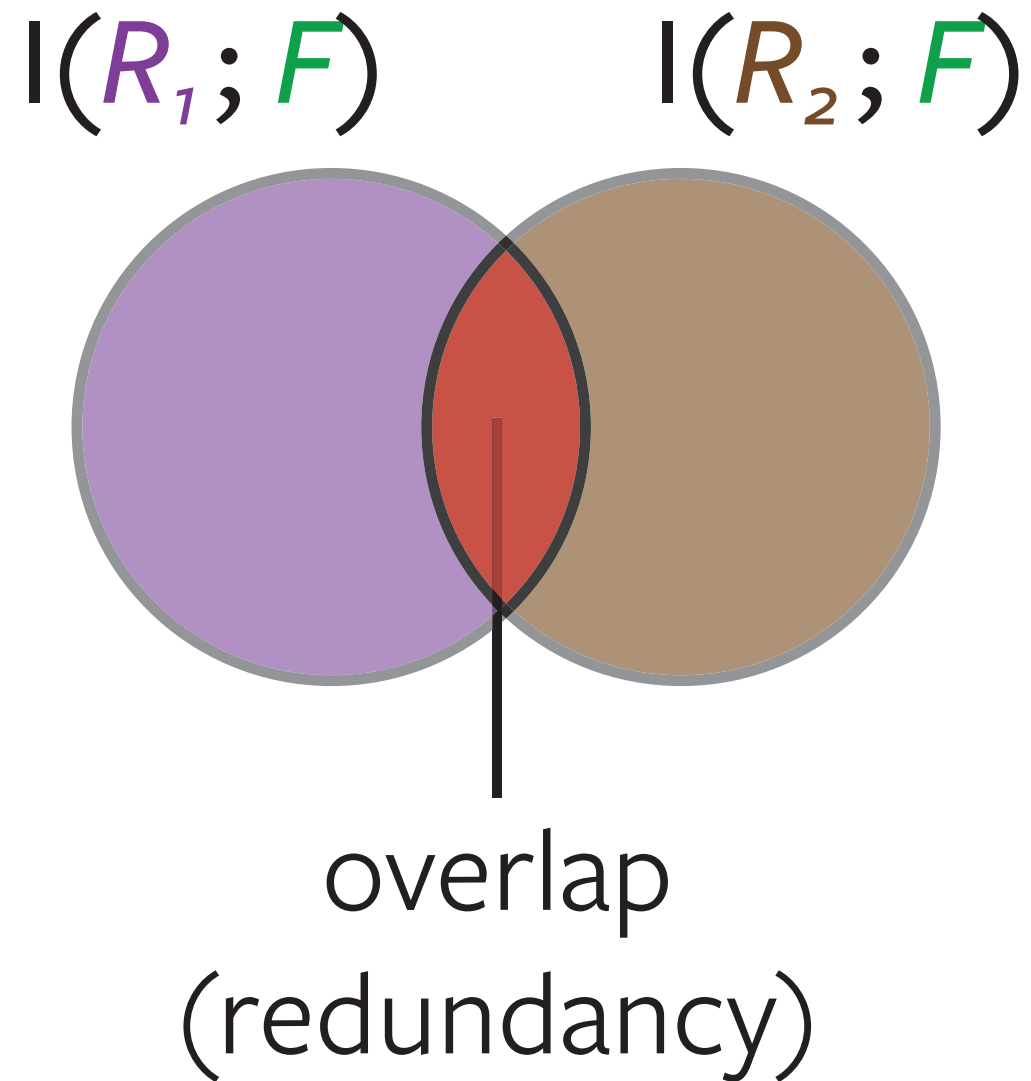


Experimental Modalities

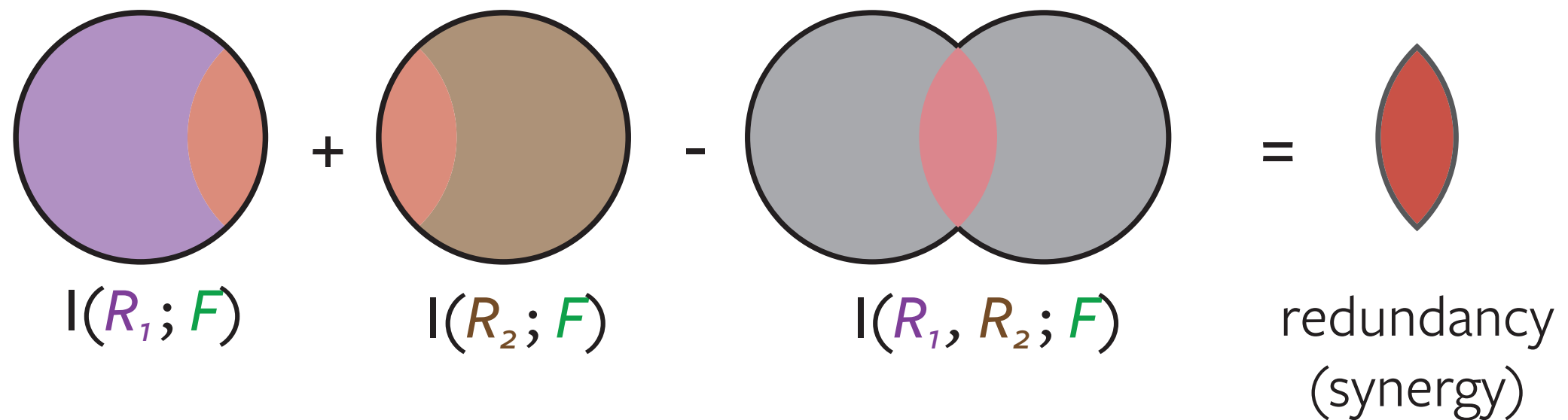
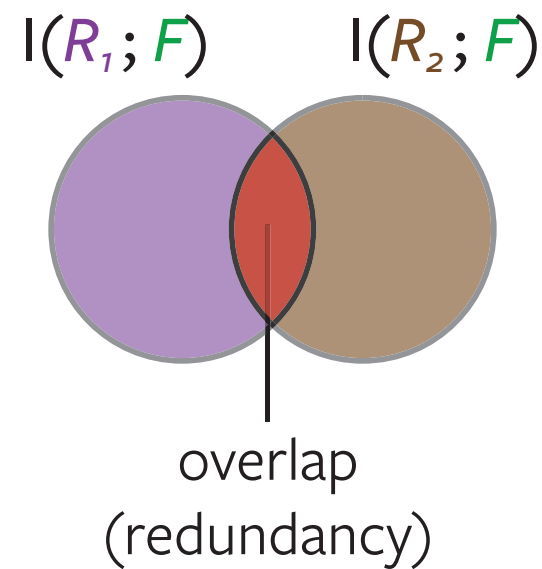


simultaneously recorded fMRI voxel activation (single trial GLM beta)

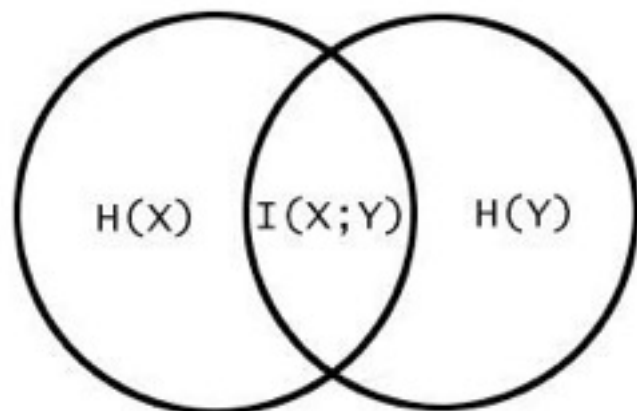
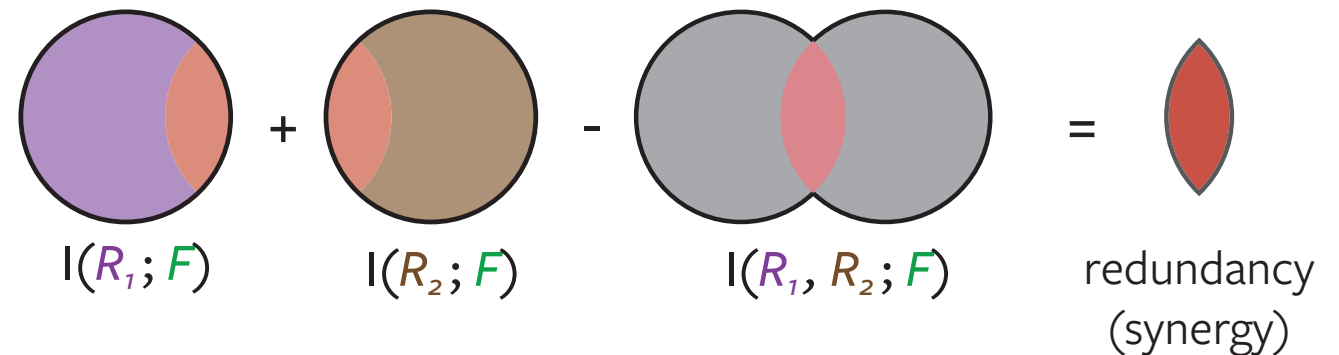
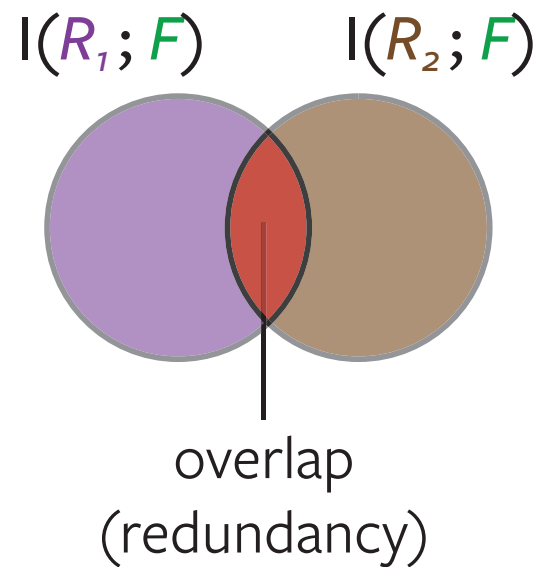
Representational Interactions



Representational Interactions



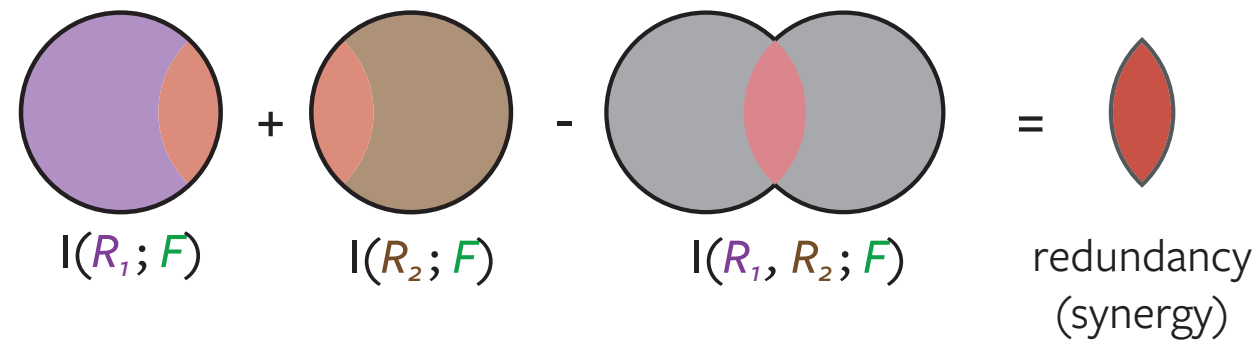
Representational Interactions



$$I(X; Y) = H(X) + H(Y) - H(X, Y)$$

- MI = overlap in entropy (shared uncertainty / variance) between two signals
- Interaction (redundancy / synergy) = overlap in MI (about an external stimulus) between two signals

Representational Interactions

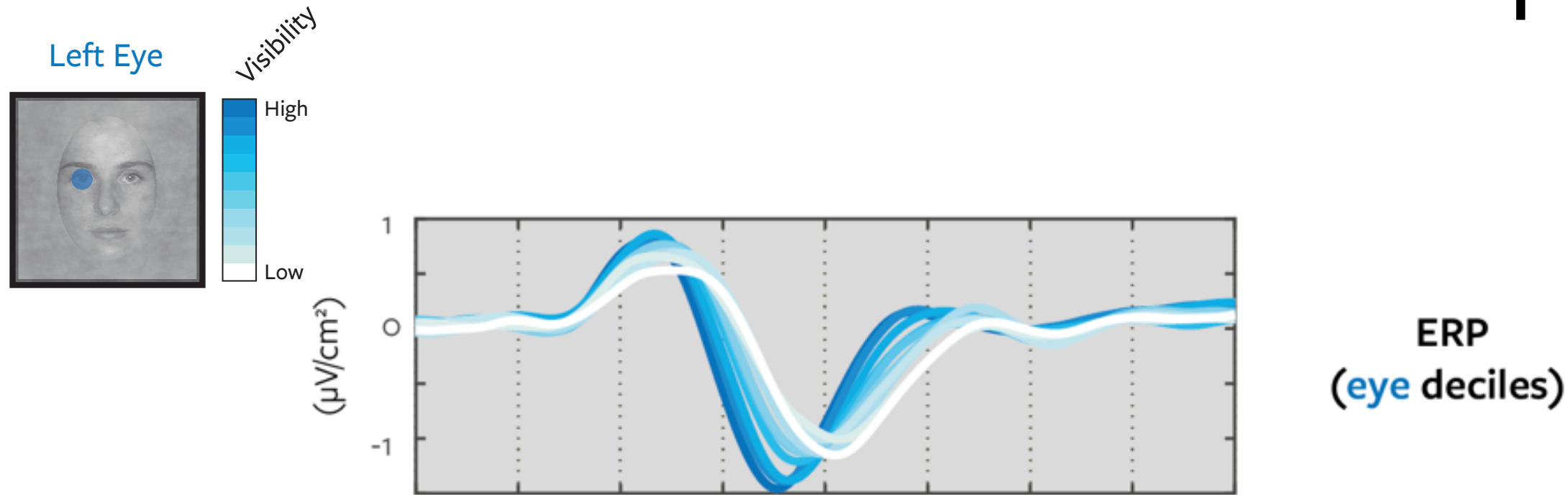


- If >0 : Representational overlap: Redundancy
- If 0 : No Overlap: Independence
- If <0 : Synergy.

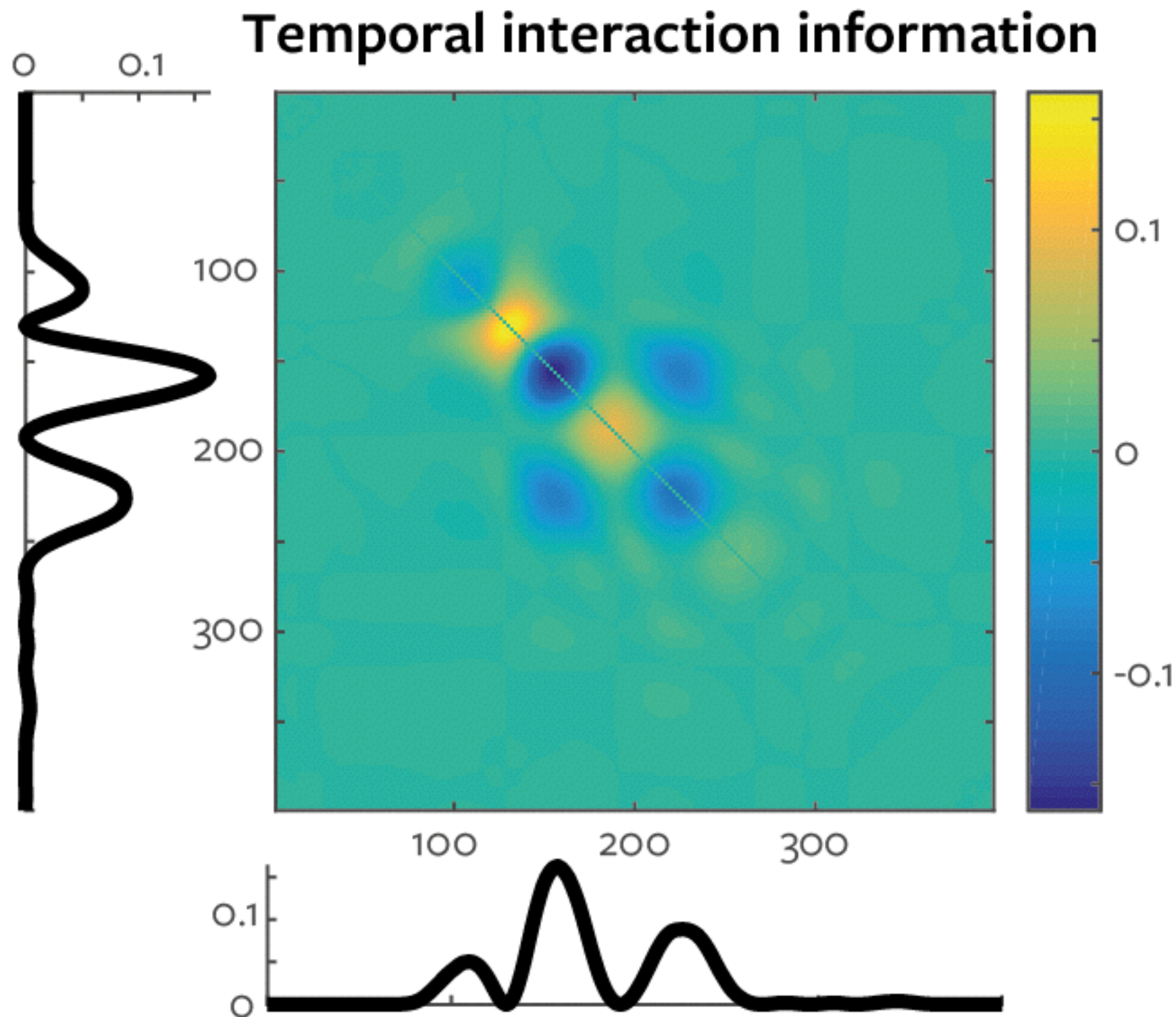
Interaction Information

- Redundancy : overlapping representation (on a trial-by-trial basis). Suggests the modulation in both responses represents the same processing mechanism
- Independence : independent representation. Suggests the modulation in the two responses reflects different processing mechanisms (different aspects on different trials)
- Synergy : trial-by-trial relationship between the signals gives extra information about the stimulus.

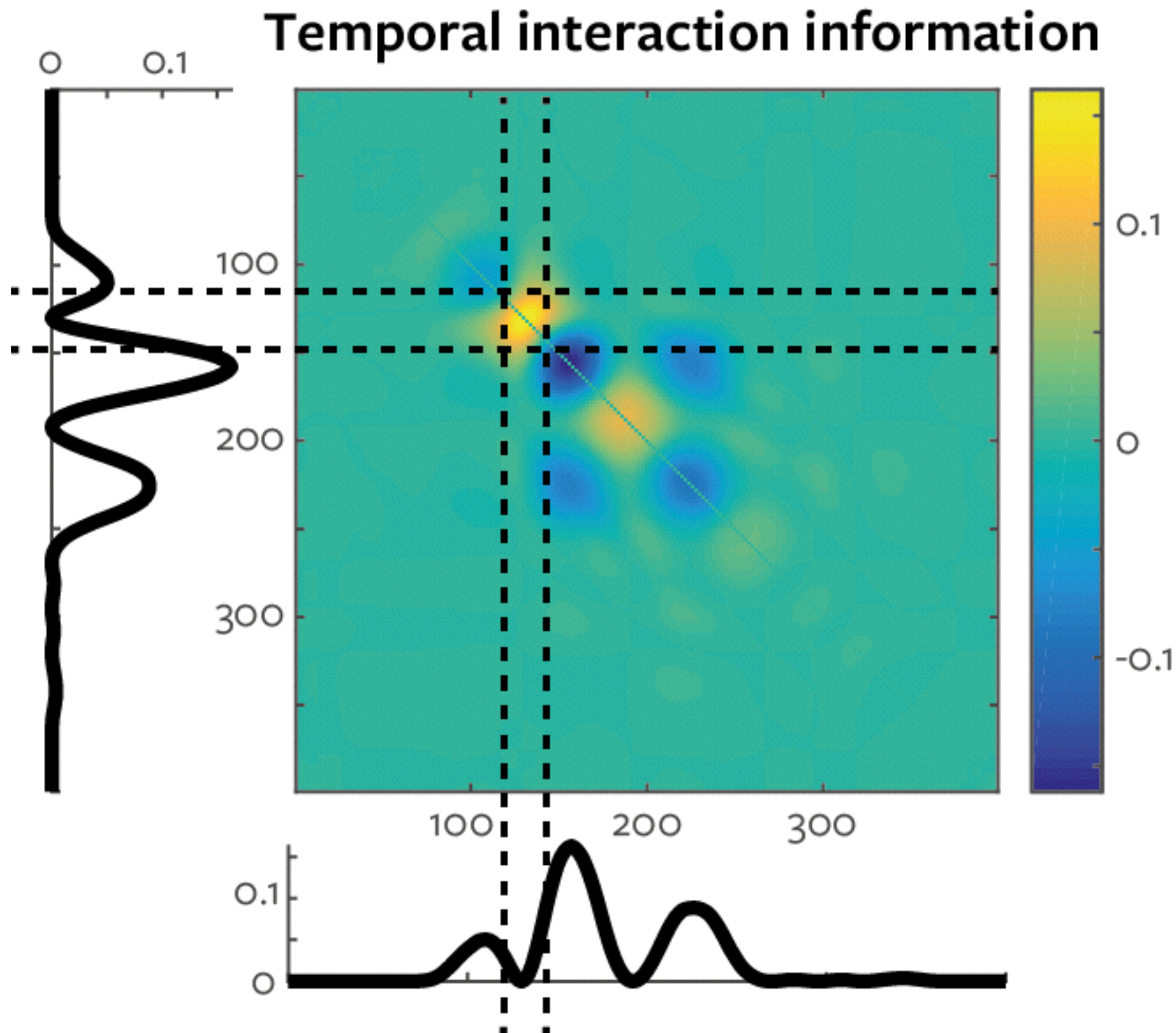
Example: event-related design, stimulus modulated evoked response



Example: temporal interaction

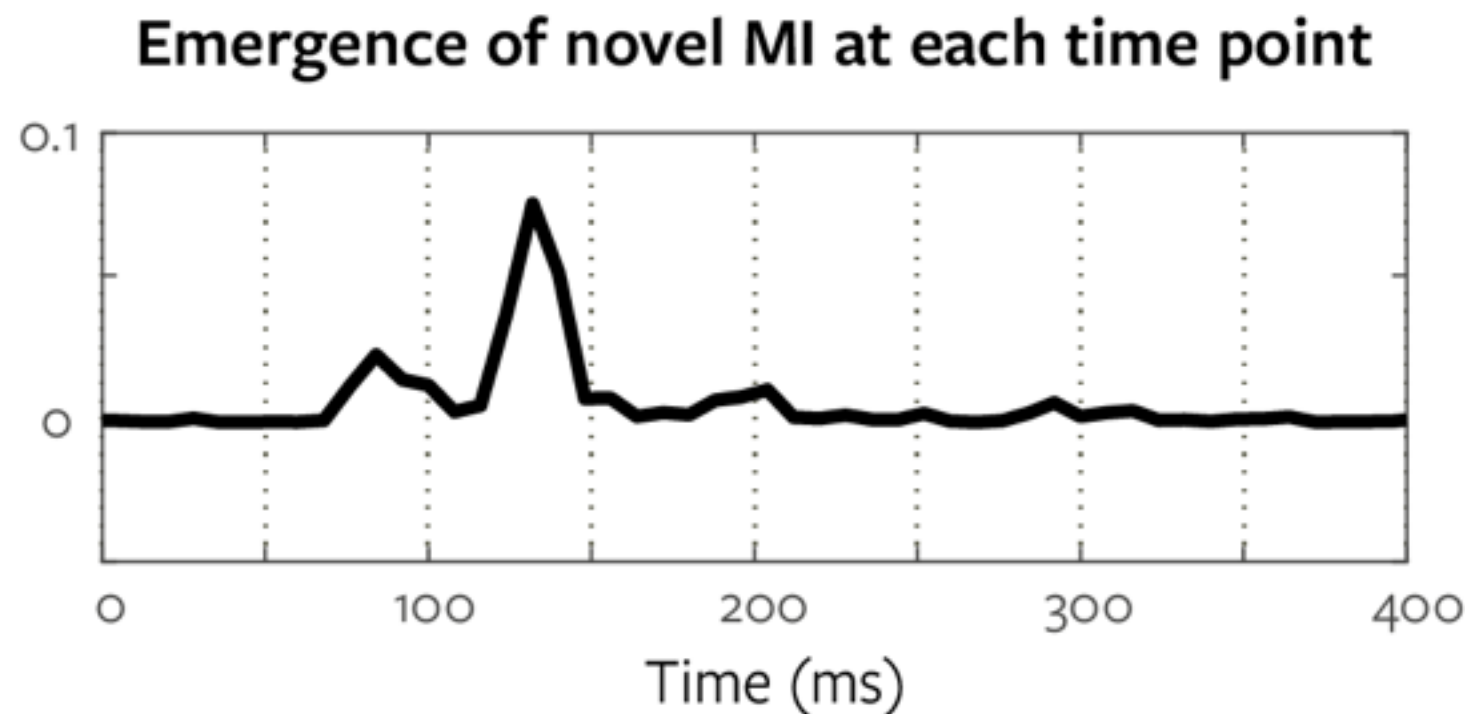


Example: temporal interaction

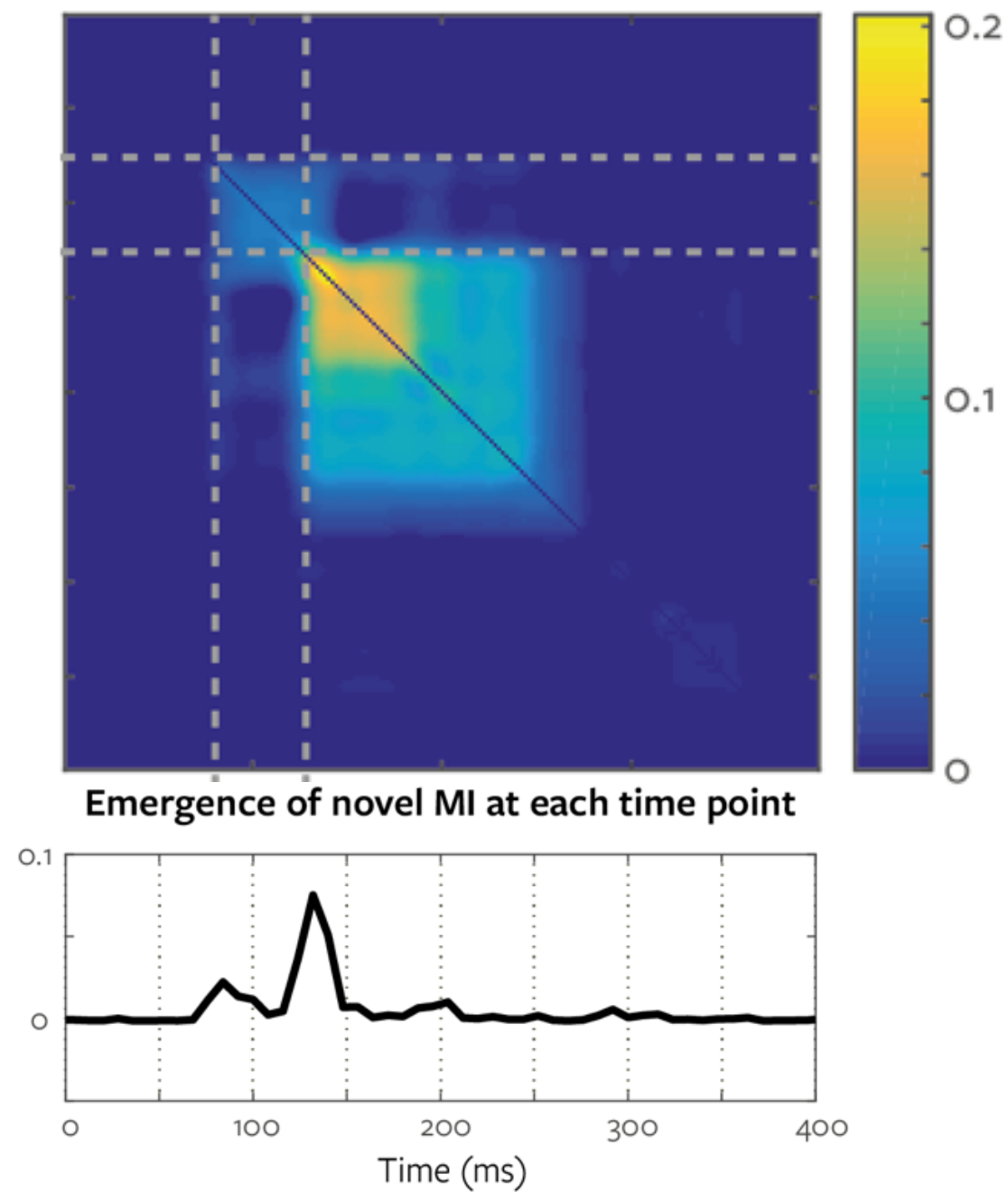


Example: novel information

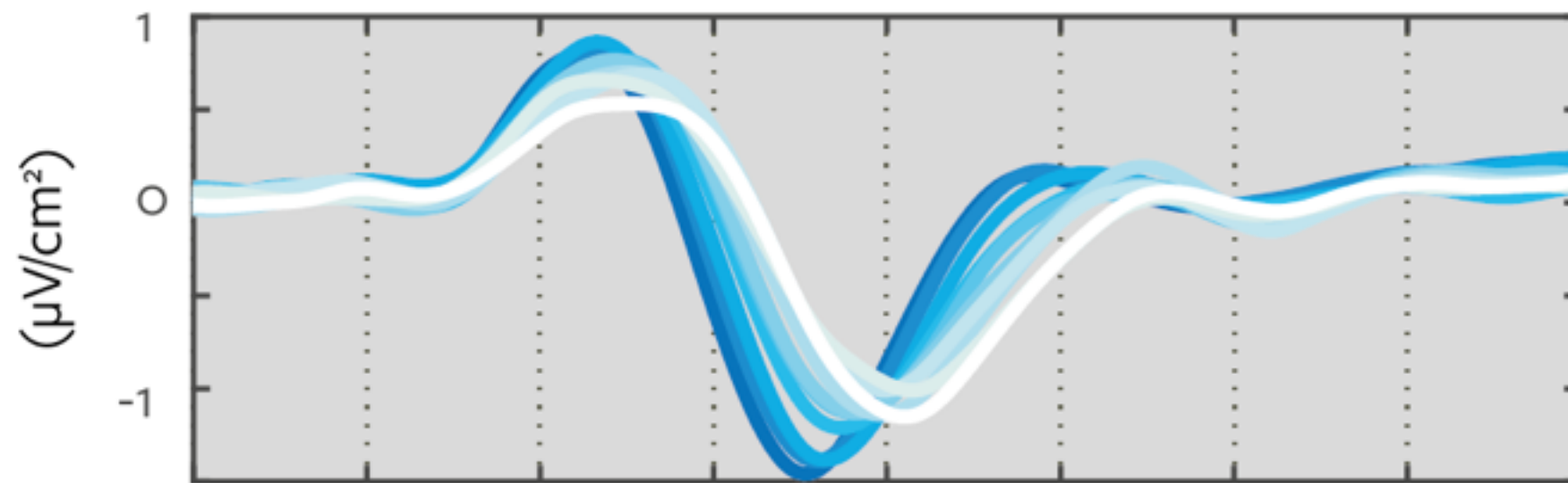
- How much information do we gain from observing EEG[t] when we already know EEG[t-1]?
- $MI(EEG[t], EEG[t-1] ; STIM) - MI(EEG[t-1] ; STIM)$
 $= CMI(EEG[t]; STIM \mid EEG[t-1])$



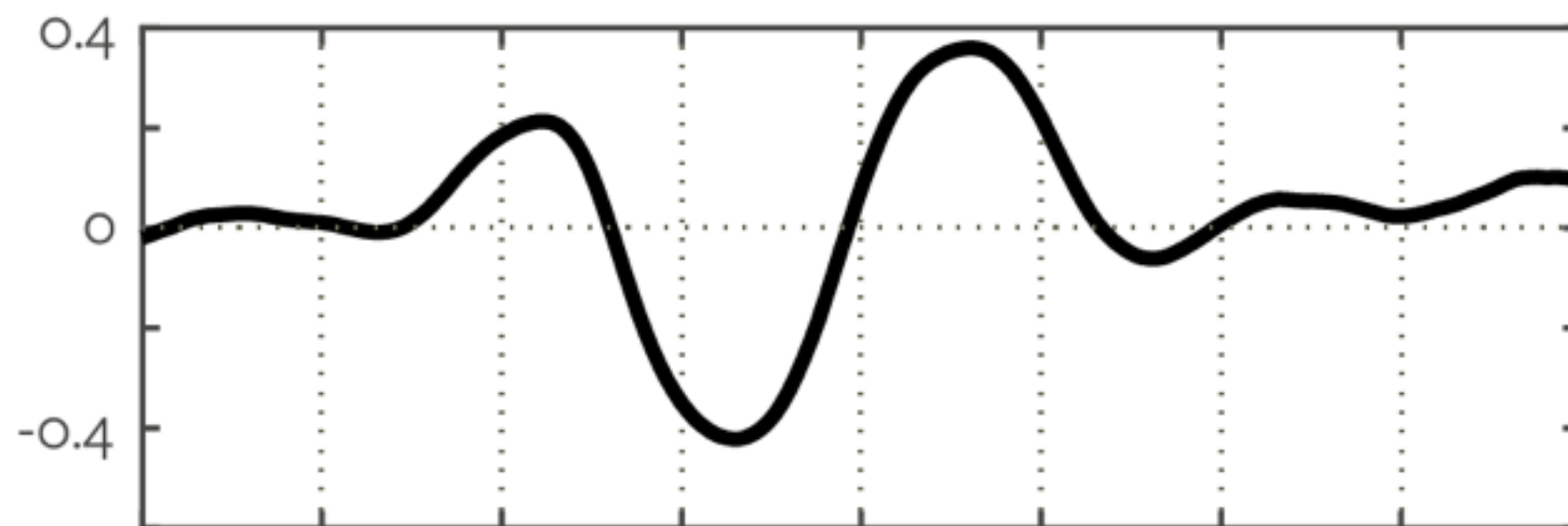
Example: novel information



Recap



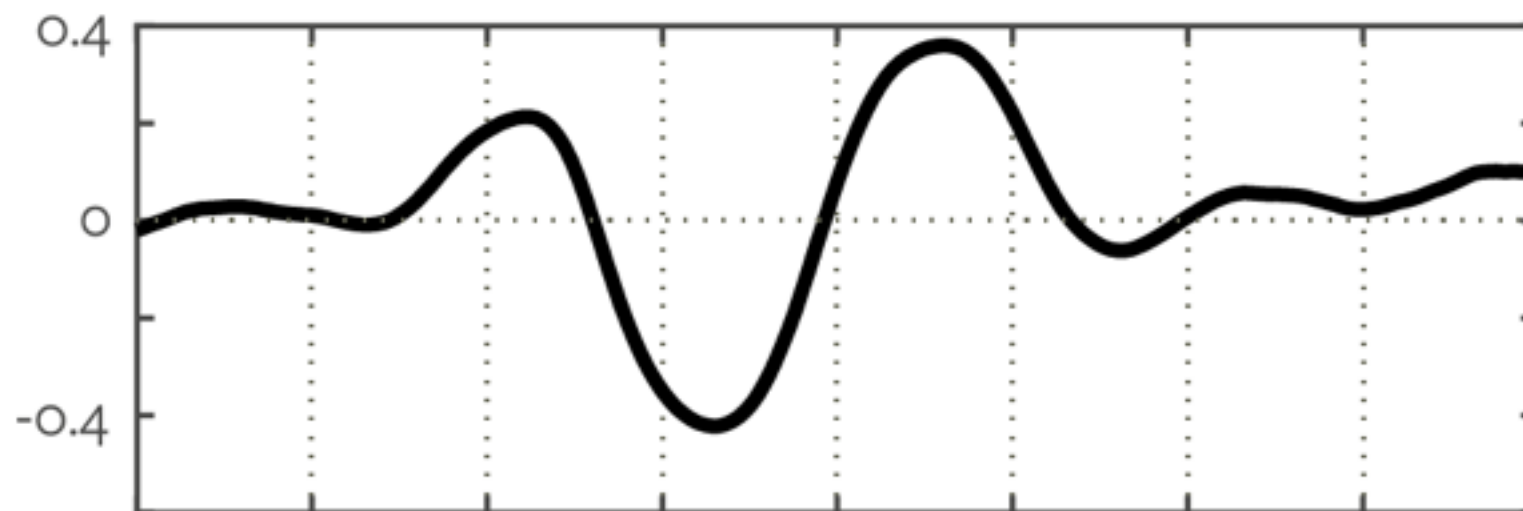
ERP
(eye deciles)



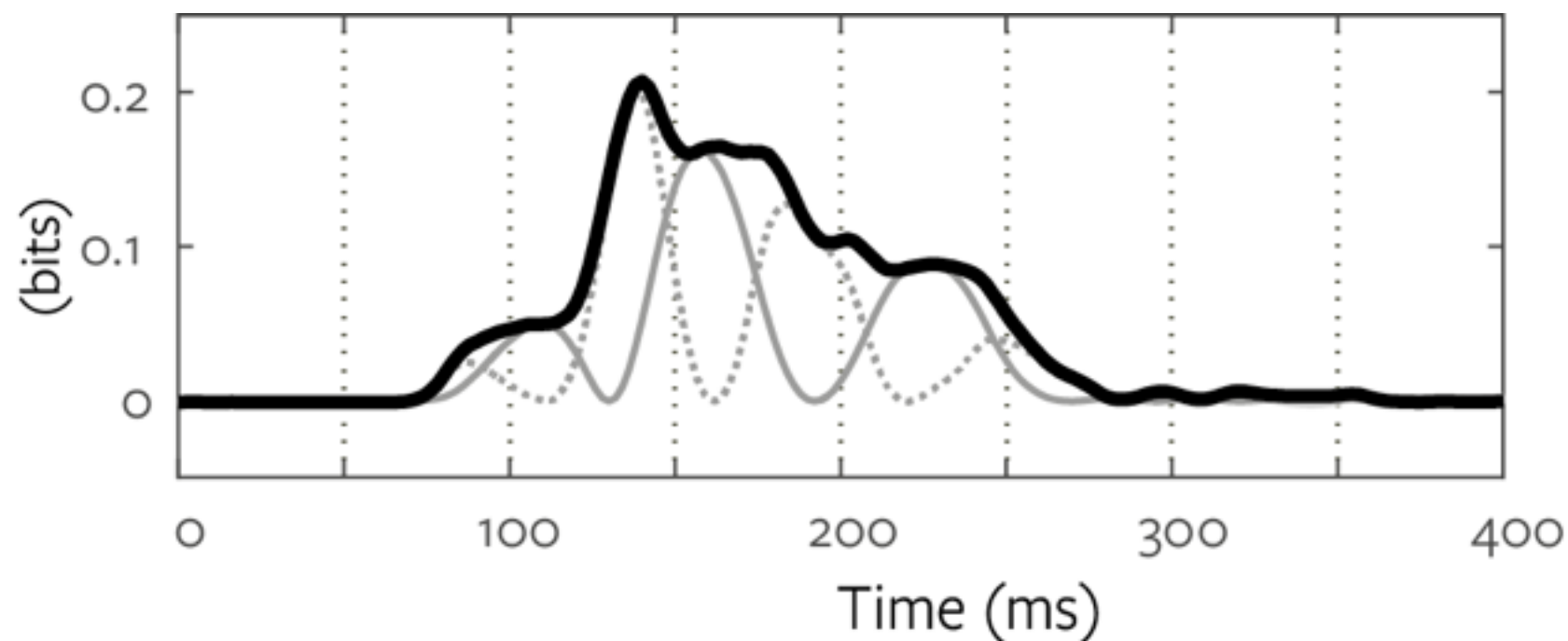
Rank Correlation
(eye, EEG)

Recap

- Where and how strongly does my experimental intervention affect my recorded responses?

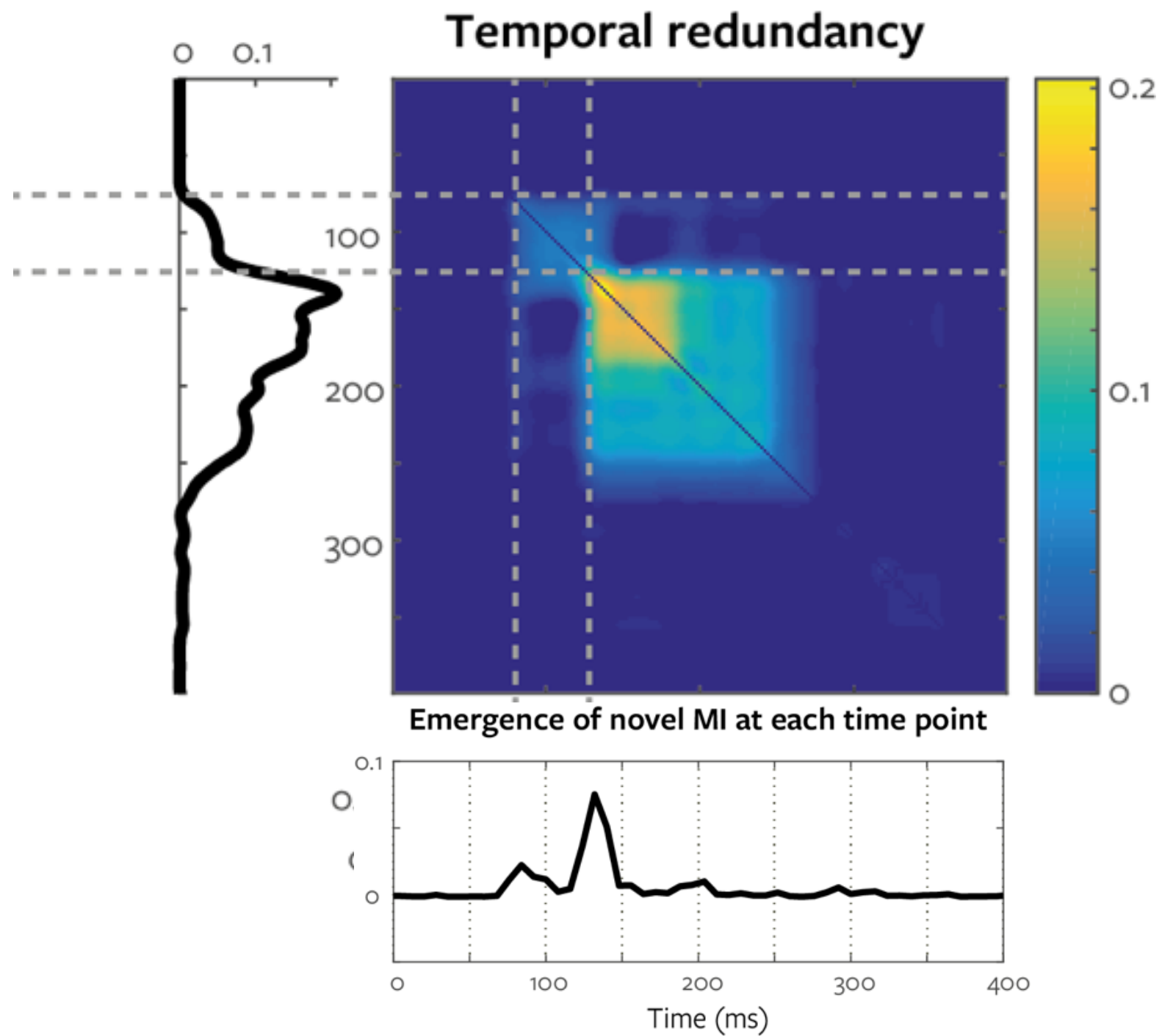


Rank Correlation
(eye, EEG)



MI(eye; [EEG dEEG])

Recap



Representational Interactions vs RSA

- Does not require high dimensional responses - allows greater temporal + spatial resolution (single sensor / time point)
- Does not require discrete exemplar stimuli (can work with dynamic naturalistic stimuli or simple contrasts)
- RSA can only detect overlap (redundancy), but info. theory approach can also identify synergy
- Can condition out other correlated features through all calculations

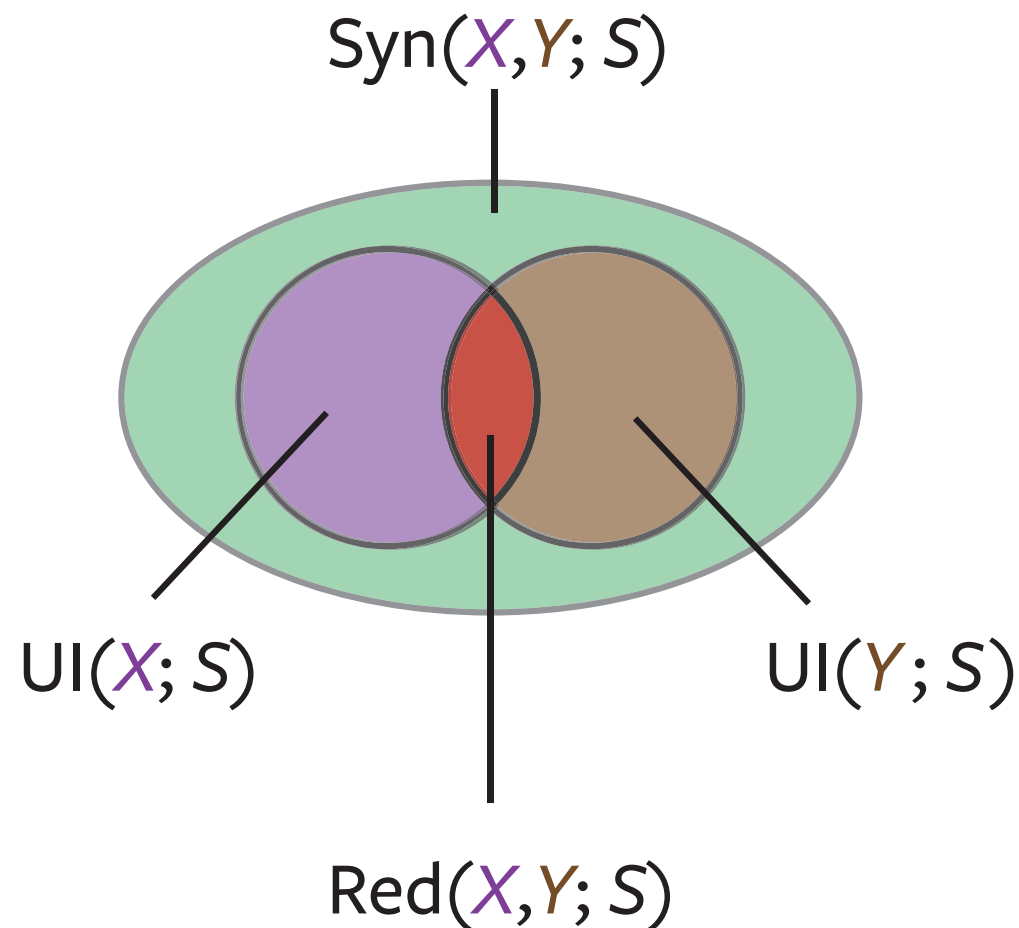
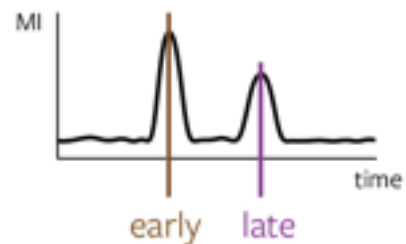
Representational Interactions vs Cross-temporal decoding

- Does not require high dimensional responses
- Temporal generalisation can only detect similar representations (not synergy)
- Temporal generalisation: what can be extracted from a form learned from the other time point. Asymmetric. Information theory - directly quantifies the shared / common change in uncertainty about the stimulus (symmetric)

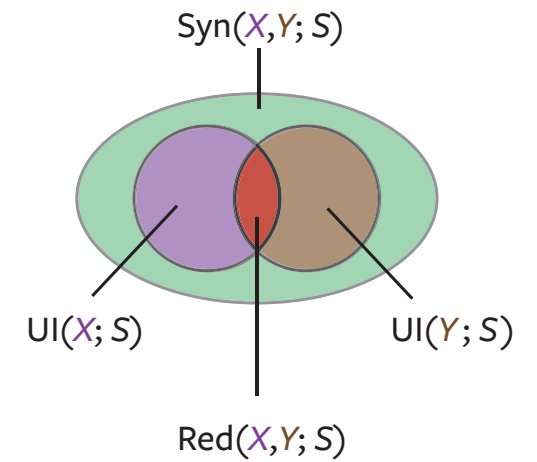
Partial Information Decomposition

- Problem: Interaction Information = Synergy - Redundancy (net effect)

stimulus modulation of evoked signal on parietal EEG electrode

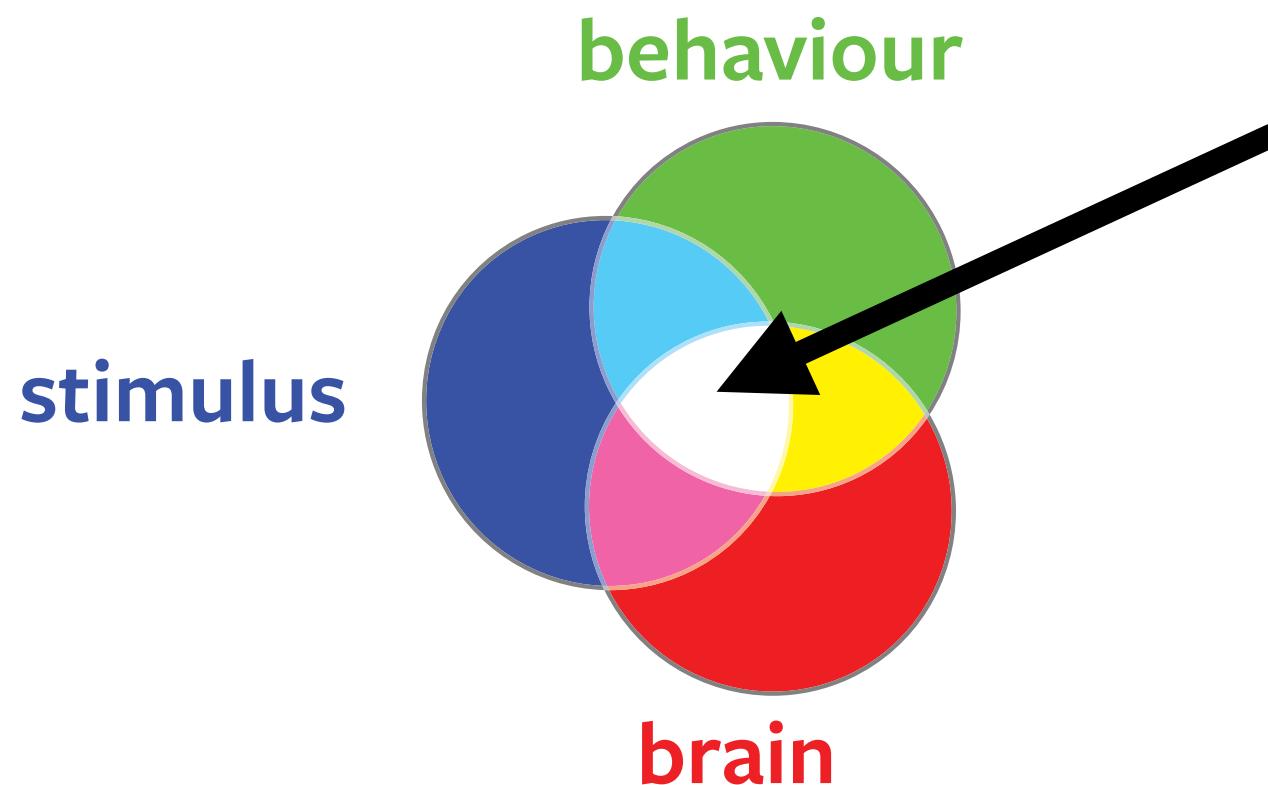


Partial Information Decomposition



- Problem: Interaction Information = Synergy - Redundancy (net effect)
- Partial Information Decomposition (Williams and Beer, 2010) provides a method to separate these contributions to the joint information.
- Depends crucially on a measure of redundancy:
Ince (2016) *Measuring multivariate redundant information with pointwise common change in surprisal*
<http://arxiv.org/abs/1602.05063>

Interactions with Behaviour



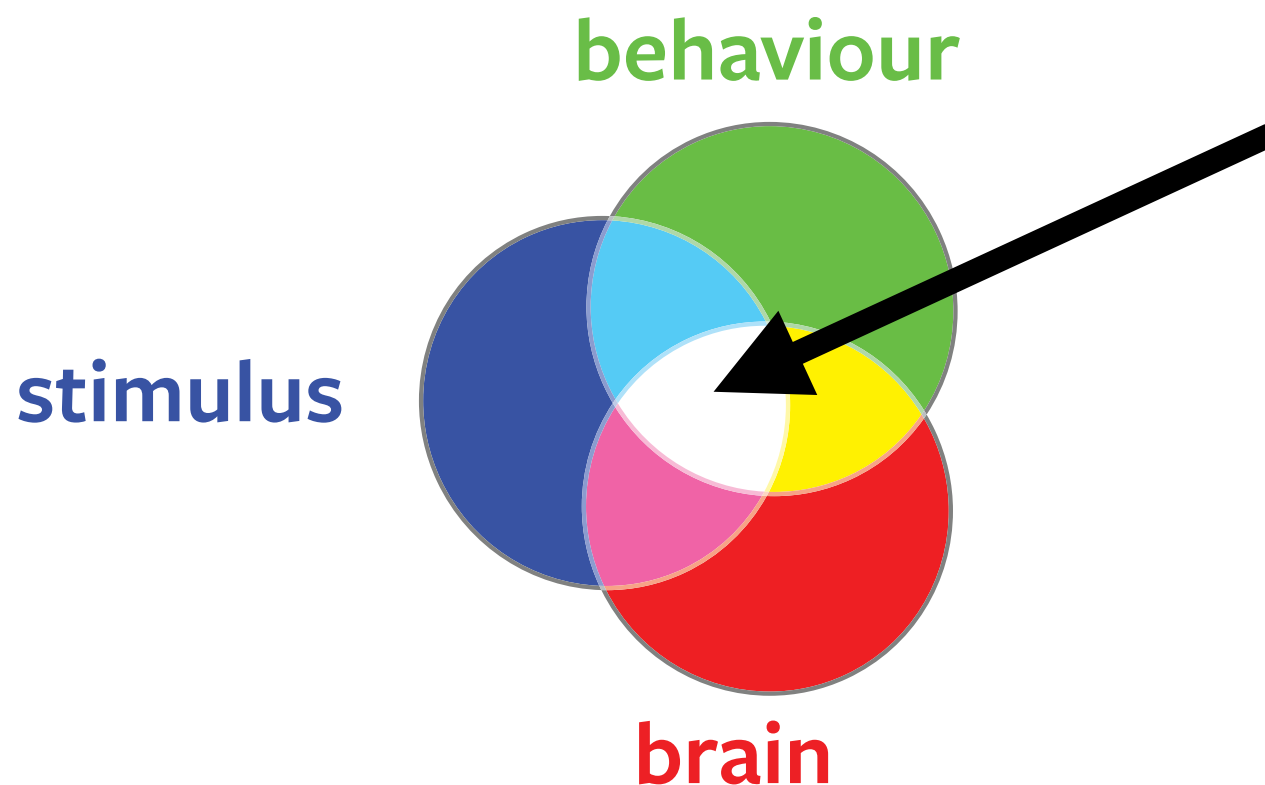
Redundancy

Stimulus variation
which commonly
effects both
behaviour and neural
signal:

task-relevant stimulus
coding

Interactions with Behaviour

Synergy



Improve prediction of
behavioural response
when stimulus +
neural signal are
considered together

decision modulated
stimulus coding

Group Stats?

- Single subject significance is a much stronger result! (ie number of subjects significant)
- Conventional group stats across subjects (treat MI quantities as the experimental measure)
- Population prevalence inference
<https://github.com/allefeld/prevalence-permutation>

Valid population inference for information-based imaging: From the second-level *t*-test to prevalence inference



Carsten Allefeld^{a,*}, Kai G6rgen^{a,1}, John-Dylan Haynes^{a,b,1}

^aBernstein Center for Computational Neuroscience, Berlin Center of Advanced Neuroimaging, Department of Neurology, and Excellence Cluster NeuroCure, Charit6 – Universit6tsmedizin Berlin, Germany

^bBerlin School of Mind and Brain and Department of Psychology, Humboldt-Universit6t zu Berlin, Germany

Summary

- A practical statistical framework for neuroimaging data analysis based on information theory
- A simple statistical function (plug in replacement for correlation) that can handle multiple different statistical comparisons (multivariate, continuous, discrete) with **effect sizes** on a meaningful (additive) common scale.
- Many related quantities that allow addressing questions that are difficult to address with classical statistics (correlated features, representational interactions, connectivity and communication).
- Examples here were event-related, but can also be used for entrainment / continuous designs

Summary (Approach)

- The brain is an organ of information processing: therefore an information processing perspective should be useful for neuroimaging analysis
- Systematic stimulus sampling
- What information is used for behaviour?
- Where/When its represented in the brain signals?
- Relationship between information content of different signals (cf RSA)

Summary

- Soon available in Fieldtrip!



```
%%  
cfg = [];  
cfg.design = data.trialinfo(:)';  
cfg.statistic = 'ft_statfun_gcmi';  
cfg.gcmi.method = 'cd_model';  
cfg.gcmi.complex = 'complex';  
cfg.precondition = 'before';  
cfg.numrandomization = 100;  
cfg.method = 'montecarlo';  
  
stat = ft_timelockstatistics(cfg, tlck);
```

Information theoretic quantity	Other statistical approaches
Mutual Information (discrete; discrete)	Chi-square test of independence; Fishers exact test
MI (univariate continuous; discrete)	2 classes: T-test, KS-test, Mann-Whitney U test; ANOVA
MI (multivariate continuous; discrete)	2 classes: Hotelling T ² -test; Decoding (CV classifier)
MI (univariate continuous; univariate continuous)	Pearson correlation; Spearman rank correlation; Kendall rank correlation
MI (multivariate continuous; univariate continuous)	Generalized Linear Model framework Decoding (CV regression)
MI (multivariate continuous; multivariate continuous)	Canonical correlation analysis Distance correlation
Conditional Mutual Information	Partial correlation (continuous variables and linear effects only)
Directed Information	Granger causality
Directed Feature Information	Dynamic Causal Modeling (Psychophysiological Interactions)
Interaction Information	Representational Similarity Analysis (redundancy only) Cross-classification decoding (redundancy only) Mediation analysis