

# Principle ERP reduction and analysis

Estimating and using principle ERP waveforms underlying ERPs  
across tasks, subjects, and electrodes

Emilie Campos, MS

Department of Biostatistics  
University of California, Los Angeles

Joint work with:

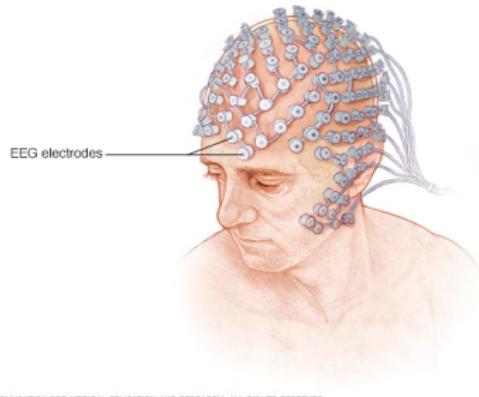
Chad Hazlett, PhD (UCLA), Patricia Tan, PhD (UCLA), Holly Truong (UCLA), Sandra Loo, PhD (UCLA), Charlotte DiStefano, PhD (UCLA), Shafali Jeste, PhD (UCLA), Damla Şentürk, PhD (UCLA)

# Outline

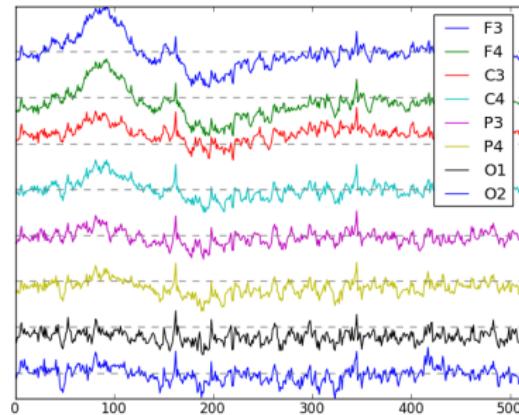
# Motivation

# Current ERP analysis

- Electroencephalogram (EEG) is a non-invasive tool to capture changes in voltage measured at the scalp
- EEG recordings time-locked to an event of interest, such as the onset of a trial or a participants response, are event-related potential (ERP) waveforms



© MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH. ALL RIGHTS RESERVED.



## Current ERP analysis: issues

- Utilizing contrasting task conditions, task-related changes to the ERP waveform are scrutinized to identify the ERP components that are thought to reflect particular brain processes
- Standard approach for measuring ERP components: take the average or peak amplitude over an investigator-selected time window, when the target ERP is expected to peak
- Overlapping activity of these unknown components can make the amplitude in a targeted interval higher or lower
- Problematic to attribute a peak that is observed at a similar time to the same component, with the same functional meaning that was described in the previous studies

## Alternatives to standard approach

- Combine Principle Components Analysis (PCA) and Independent Component Analysis (ICA)
- Early works: used ICA to carry out subject-level decompositions of the signal, generally relying on clustering (Makeig et al., 1996)
- Multi-subject decomposition methods:
  - Spatiotemporal PCA (Spencer et al., 2001) – two PCA steps: a spatial one that reduces the electrode dimension across subjects, and a temporal PCA
  - Multi-level group ICA (mIGICA) (Eichele et al., 2011) and temporal-concatenation group ICA (tcGICA) (Cong et al., 2013) – both consider trial-wise ERP and conduct one or two PCA steps for dimension reduction at the electrode or subject levels, followed by a final ICA step for source separation

## pERP-RED algorithm

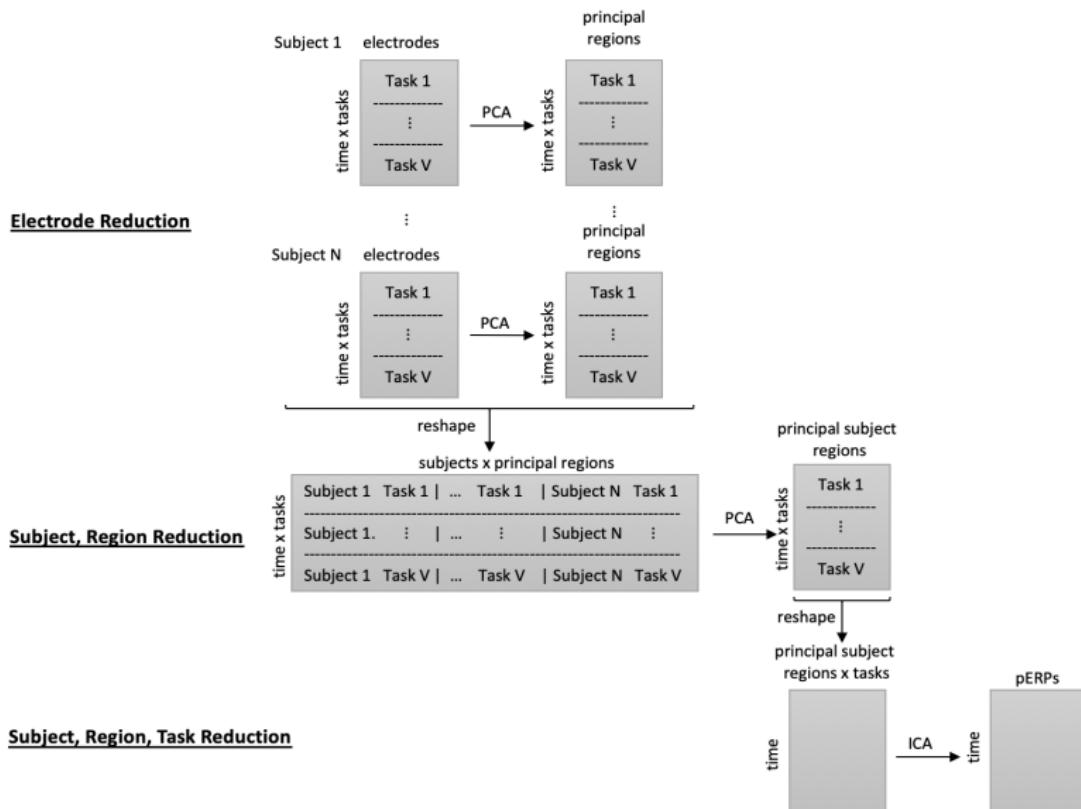
# How is pERP-RED better

- Similar in spirit to other multi-subject approaches
  - First PCA: reduce electrodes separately for each subject
  - Second PCA: reduce the subject-region-task specific ERP averages into a smaller set that explains most of the sample variation
  - Followed by an ICA step for blind source separation
- Designed for reducing data not only across multiple subjects but also multiple tasks
- Electrode dimension reduction PCA steps of spatial PCA, mIGICA and tcGICA assume no missingness on electrodes, identical trial orderings, identical scalp topographies, or identical projections of components onto electrodes across subjects
  - Electrode reduction PCA step of pERP-RED avoids these assumptions by running dimension reduction separately for each subject in the first step

- Approach to ERP analysis that avoids using peak/mean amplitude
- ERPs are assumed to be a weighted combination of underlying signals, called pERPs
- An accessible approach to analyzing ERPs in terms of their pERPs
- A set of tools, called “pERP-space analysis”
- Easy-to-use software to conduct these analyses in R

- Motivated by the goal of estimating an underlying set of component waveforms, herein referred to as pERPs
- ERPs formed by time-locked averages at any given electrode, participant, or trial type is approximately a weighted combination of these pERPs
- Involves (i) a series of data concentrating steps that turn a larger number of noisy waveform records into a smaller number of less noisy ones and (ii) steps that generate maximally independent, unmixed components from these concentrated signals
- Notation
  - $i = 1, \dots, N$ : subjects
  - $v = 1, \dots, V$ : tasks
  - $e = 1, \dots, E_i$ : electrodes
  - $t = 1, \dots, T$ : time points
  - $p = 1, \dots, P$ : pERPs

# pERP-RED: Algorithm schema



# pERP-RED: Algorithm

- ① *Data initialization.* Data are split into a training and test set.  
Normalize each of the ERPs to unit variance.
- ② *Electrode reduction.* Apply PCA to the  $N$  subject-specific matrices.
- ③ *Subject-region reduction.* Reshape the data above into a matrix with all of the principal regions as the columns and the task and time data concatenated in the rows. Apply PCA to generate  $N_R$  principal subject-regions.
- ④ *Source separation.* Reshape data into a matrix with all task principal subject-regions as the columns and time as the rows. Fast ICA is then used to produce  $P$  principle ERPs, where  $P$  may be chosen by regressing the true signal onto the pERPs and obtaining an  $R_{\text{test}}^2$  value.

- Use subject-level averages over trials of a given type as the “records” that first enter the algorithm – need not be the case
  - i.e., interested in practice effects → earlier trials and later trials may be averaged separately
- Splitting into training and test set: choose the appropriate dimensionality of the data in the final step → pERPs re-estimated using all of the data
- Normalize the records in each data reduction step: covariance matrix is actually a correlation matrix
  - Electrode reduction: go from a large number of electrodes that may have highly correlated signals to a smaller set, we call “regions”, each of which provides uncorrelated information
  - Subject-region reduction: if there are groups of participants that have “region” signals that are highly correlated, this information can be combined with little loss

- Proportion of variance each of the PCA steps must explain during the data concentration steps
  - Keeping a larger number of components implies that more of the data will survive but at the cost of keeping more noise
  - Our default: choosing to keep enough components to cover 80% of the variation is that this should be sufficient to recover almost all of the true signal of value
- Choosing  $P$ 
  - $R_{\text{test}}^2$ : the proportion of variation in the test set explained by the estimated pERPs
  - Choose the number of pERPs  $P$  according to how well the set of estimated pERPs can explain the test data
  - Choose the smallest value of  $P$  such that raising  $P$  would result in little gain in  $R_{\text{test}}^2$

## pERP-Space analysis

- Central concept: any observed ERP can be recast as a vector of coefficients describing the magnitude of each pERP's contribution to that ERP
  - Step 1: Individual scoring
  - Step 2: Summary across individuals
  - Step 3: Description and inference

## pERP-Space analysis: Individual scoring

- Condition  $c$ : trial type within a single experiment, i.e. match vs. mismatch in the following data application
- Regress the observed ERP denoted by the vector  $Y_{i,c,e}$  on the estimated pERPs denoted by the matrix  $\Phi$

$$\omega_{i,c,e} = (\Phi^T \Phi)^{-1} \Phi^T Y_{i,c,e}$$

- $\omega_{i,c,e}$ : the vector of scores describing the magnitude of each pERP's contribution to the ERP for individual  $i$  and condition  $c$  at electrode  $e$
- Compare condition  $c$  to condition  $c'$  by subtraction

$$\omega_{i,c-c',e} = \omega_{i,c,e} - \omega_{i,c',e}$$

## pERP-Space analysis: Summary across individuals

- Let  $G_g$  denote the set of subject indices in group  $g$  and  $N_g$  denote the size of the group
- Group mean: average pERP contribution for the group

$$\bar{\omega}(g)_{c-c',e} = \frac{1}{N_g} \sum_{i \in G_g} \omega_{i,c-c',e}$$

- Across participant standard deviation (APSD): variability in the loadings across individuals in group  $g$

$$APSD(g)_{c-c',e} = \sqrt{\sum_{i \in G_g} \frac{(\omega_{i,c-c',e} - \bar{\omega}(g)_{c-c',e})^2}{N_g - 1}}$$

- Standard errors for inference

$$SE(g)_{c-c',e} = \sqrt{\frac{\sum_{i \in G_g} \frac{(\omega_{i,c-c',e} - \bar{\omega}(g)_{c-c',e})^2}{N_g - 1}}{N_g}}$$

# pERP-Space analysis: Description and inference

- Within group condition contrast: test how much each pERP contributed to a given condition ( $c$ ) or contrast ( $c - c'$ )

$$t(g)_{c-c',e} = \frac{\bar{\omega}(g)_{c-c',e}}{SE(g)_{c-c',e}}$$

- Between group condition contrast: test how a pERP contributes to a condition or contrast for group  $g$  compared to group  $g'$

$$t(g, g')_{c-c',e} = \frac{\bar{\omega}(g)_{c-c',e} - \bar{\omega}(g')_{c-c',e}}{\sqrt{SE(g)_{c-c',e}^2 + SE(g')_{c-c',e}^2}}$$

- Determine whether certain groups have higher/lower variability than others by comparing  $\text{APSD}(g)_{c-c',e}$  across groups
- Topographic head maps can be created to show spatial distribution

- Compare  $\omega_{i,c-c',e}$  to behavior or clinical measures for person  $i$
- ERP “cleaning”: reconstruct ERPs using only the derived pERPs, leave out components deemed to be noise
- Participant rejection: pERPs explain less of an individual's signal suggests problems in data collection or other issues
- Outlier detection: individuals with very unusual values of  $\omega_{i,c-c',e}$  could be identified

# Applications

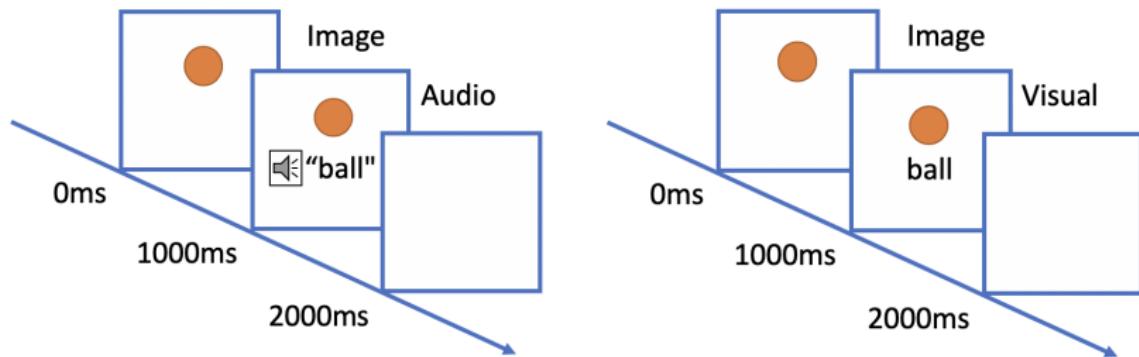
# Autism Spectrum Disorder (ASD) Study

## Application to ASD data: Study cohort

- Study cohort: 31 children aged 5-11 years old were recruited; 14 typically developing (TD), 10 verbal ASD (vASD), and 7 minimally verbal ASD (mvASD)
- Goal: study the neural mechanisms underlying language impairment in children with ASD (DiStefano, 2019)
- Diagnoses made prior to enrollment and confirmed using the Autism Diagnostic Observation Schedule (ADOS) and Social Communication Questionnaire

# Application to ASD data: Experiments

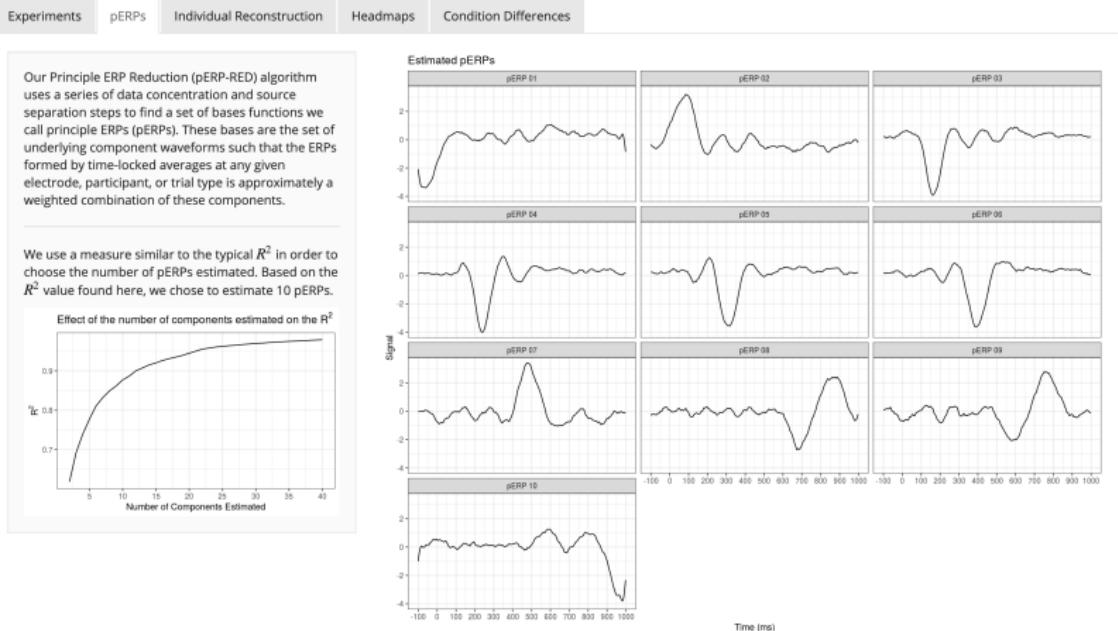
- Audio paradigm: a picture was presented and an audio recording of the spoken word was played that either matched or did not match
- Visual paradigm: a picture was presented and an image of the word appeared that either matched or did not match
- Vocabulary included 60 basic nouns (e.g., bird, dog, bike)



# Application to ASD data: Estimated pERPs

- Visit the interactive Shiny app to reproduce these results and others: [https://perpred.shinyapps.io/asd\\_exploration](https://perpred.shinyapps.io/asd_exploration)
- The  $R^2_{\text{test}}$  plot on the left and the estimated pERPs on the right

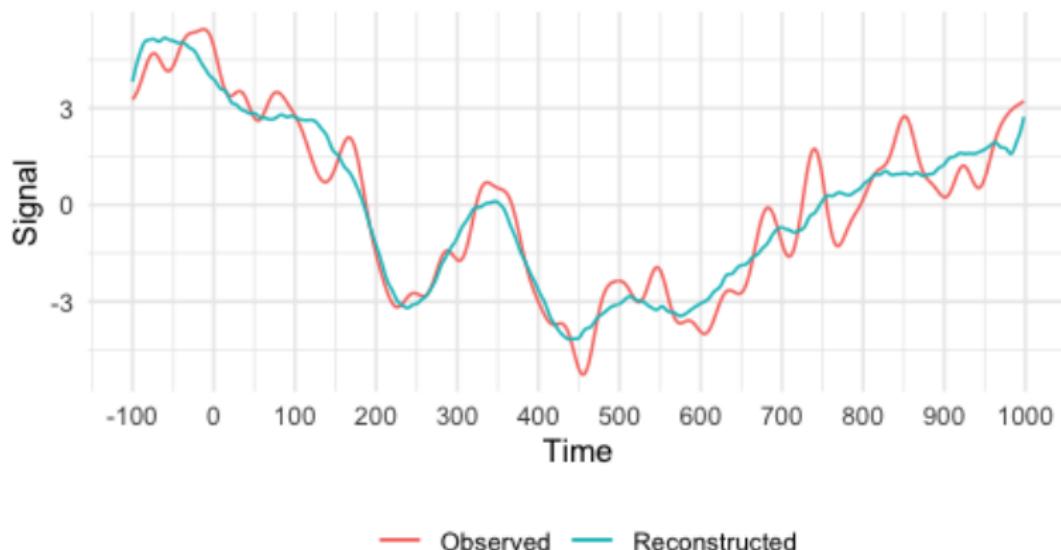
## ASD Data Exploration



# Application to ASD data: Individual reconstruction

- Individual ERPs can be reconstructed using the estimated pERPs to reduce noise

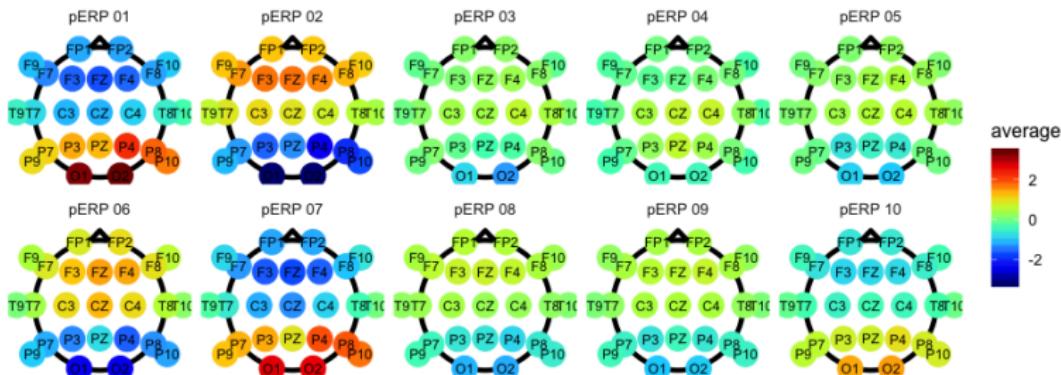
Subject 0001, Image Task, Electrode C3



# Application to ASD data: Spatial distribution

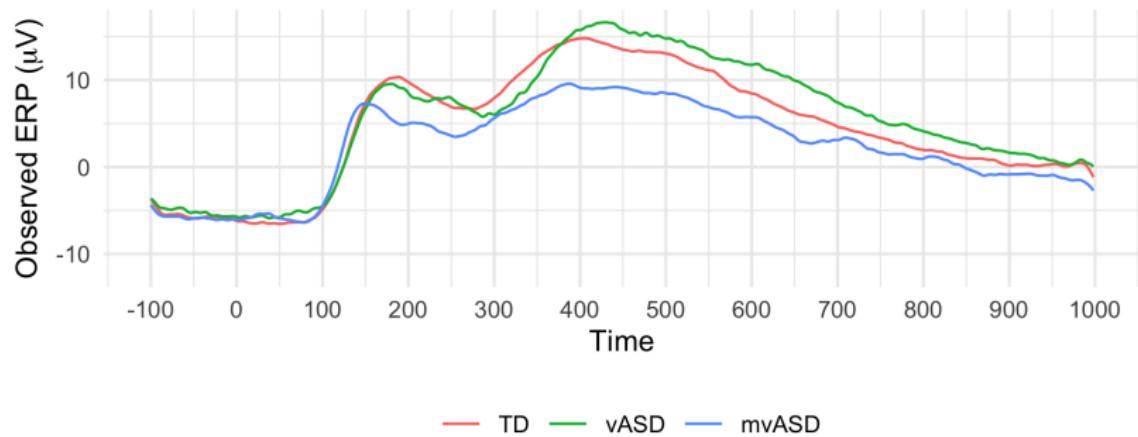
- To investigate the spatial distribution of the pERP loadings, headmaps can be used to plot the estimated coefficients
- The first two pERPs are loaded heavily onto in the O1 and O2 electrodes for the Image task

Image



# Application to ASD data: Observed image contrast

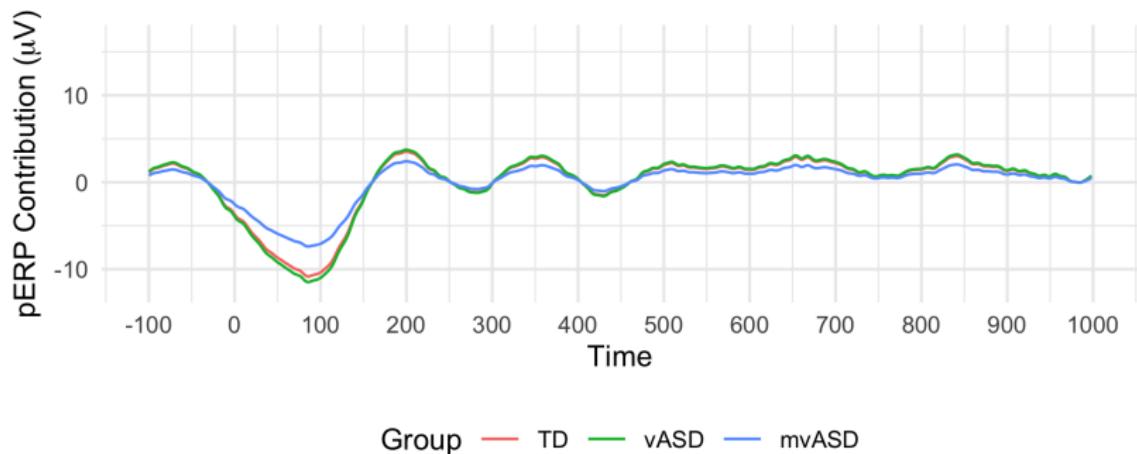
- Observed ERP for image condition at O1



- N1 is the expected reaction to a visual stimuli but isn't seen in the observed ERP, due to temporal overlap given the fast rate of trials

# Application to ASD data: pERP image contrast

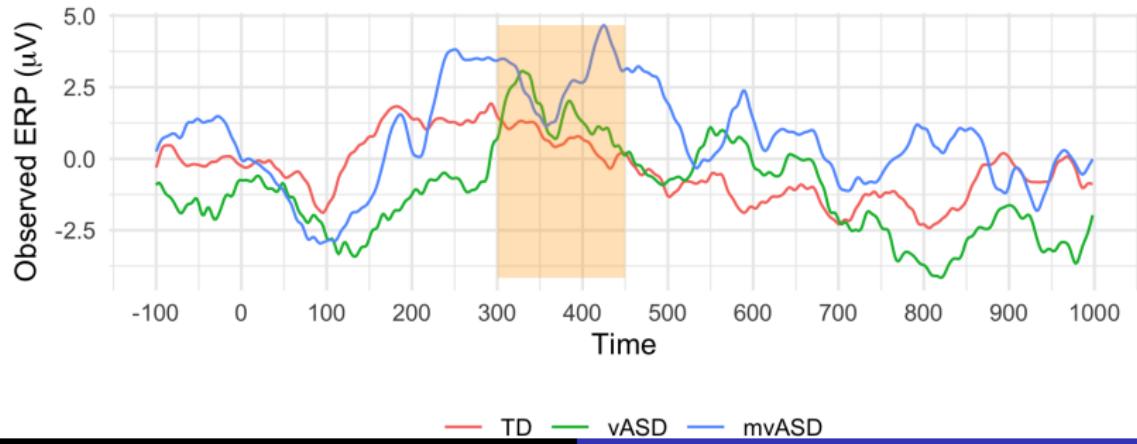
- Contribution of pERP 2 for image condition, O1



- pERP 2 contribution to the average ERP shows the expected N1, is significant in all diagnostic groups, and does not show significant group differences

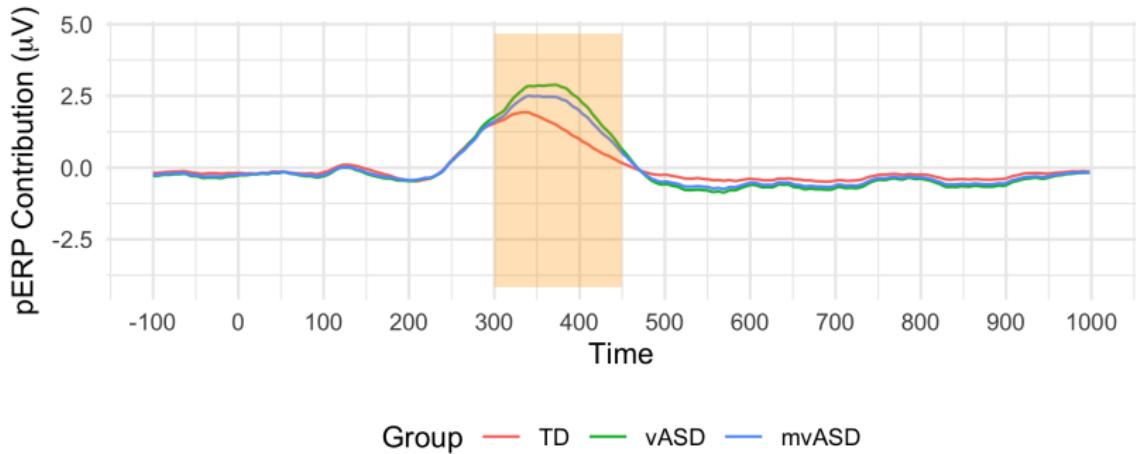
## Application to ASD data: Observed sound N4

- Main findings of DiStefano: an N4 related to primary semantic processing that was deeper for mismatch than match
- N4 used as a biomarker to assess receptive language in individuals with limited speech
- Negativity was not seen when averaged over groups, but was seen on the individual level
- Observed sound match - mismatch ERP at Pz



# Application to ASD data: pERP 5+6 sound N4

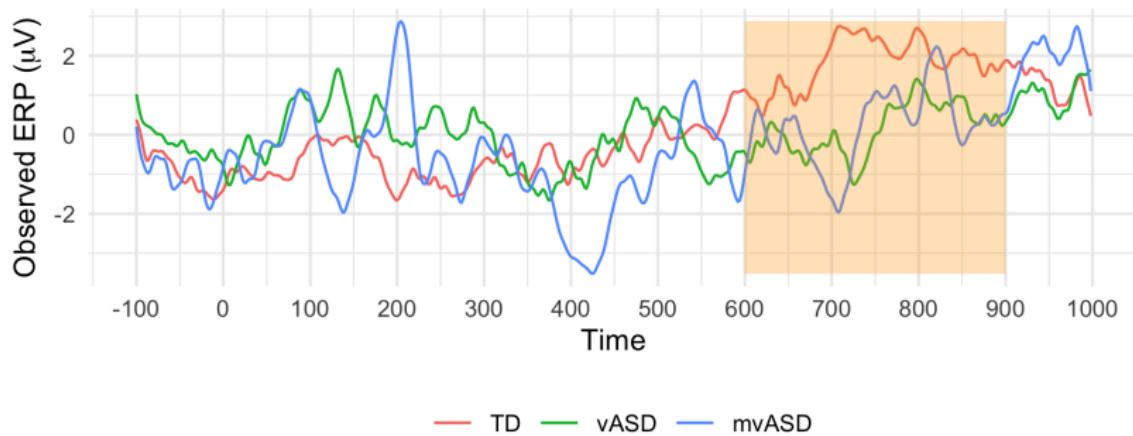
- pERP 5+6 contribution to Sound match - mismatch at Pz



- Significant within group difference at electrode Pz in pERPs 5 and 6 when comparing the contrast of match vs mismatch conditions
- Interpretation: all groups were surprised by the mismatched word

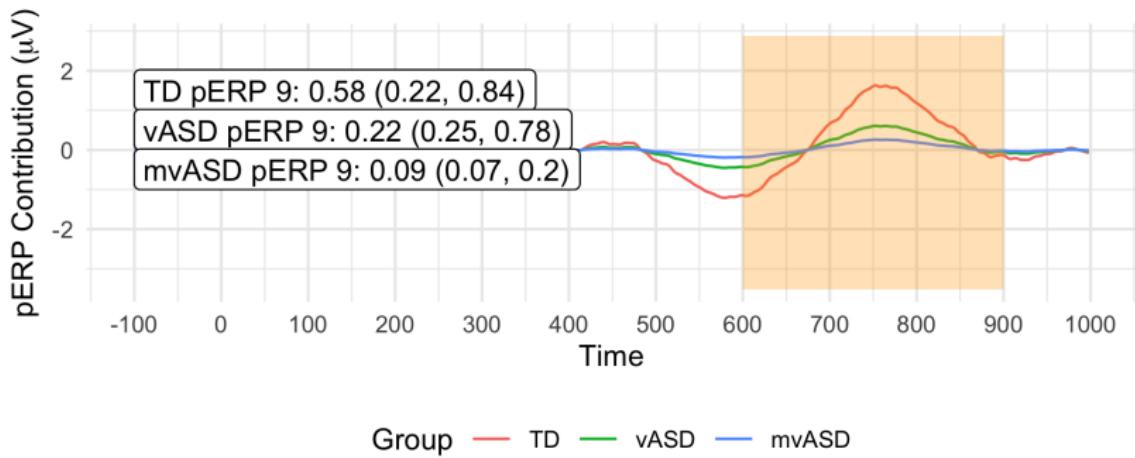
## Application to ASD data: Observed sound N600-900

- Other finding: a deeper negativity for mismatch than match trials from 600-900ms, linked to semantic integration
- Subtracting match - mismatch should lead to a long, positive peak from 600-900ms
- Sound match - mismatch observed ERP mvASD vs TD at F4



## Application to ASD data: pERP 9 sound N600-900

- pERP 9 contribution to Sound match - mismatch, plotted with mean (SD, APSD)



- A significant group difference at electrode F4 in pERP 9 when comparing TD and mvASD in the contrast of match vs mismatch
- Interpretation: while all groups were surprised by the mismatched word, only the TD group was able to integrate that information

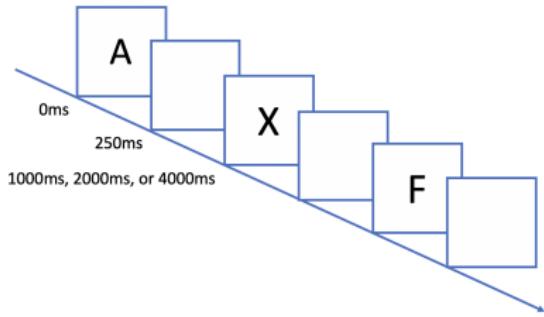
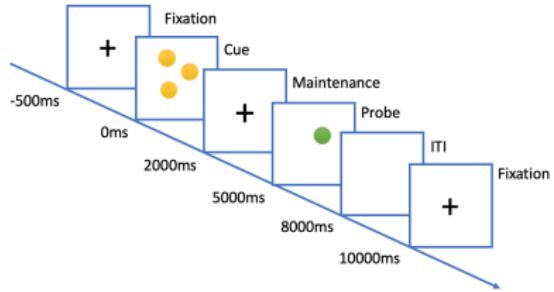
# Attention Deficit Hyperactivity Disorder (ADHD) Study

## Application to ADHD data: Study cohort

- 331 youth aged 7-17 years old, 242 with ADHD
- Goal: study of cognitive control and working memory in youth  
(clinicaltrials.gov ID: NCT00429273)
- DSM-IV diagnoses obtained through a semi-structured diagnostic interview with the primary caretaker and a direct interview

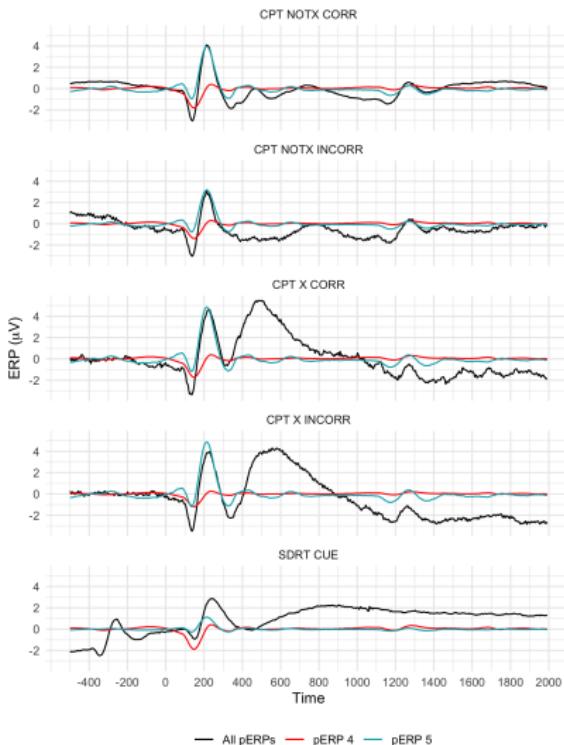
# Application to ADHD: Experiments

- Spatial delayed response task (SDRT): pay attention to the location of the yellow dots on the screen and determine if the following green dot matched the position of any of the yellow dots
- Continuous performance task (CPT): presented single letters, press and release spacebar as quickly as possible after viewing each letter
  - except when the letter is 'X'



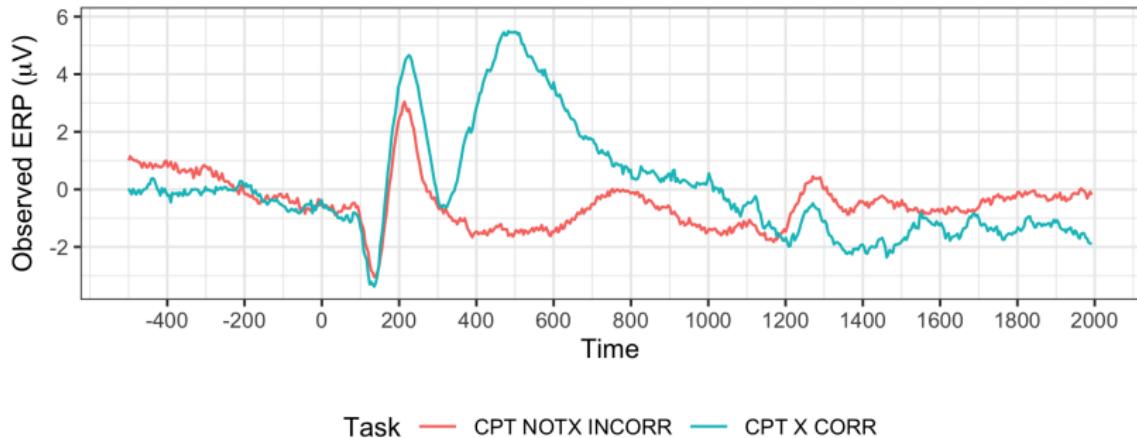
# Application to ADHD data: N1/P2 complex

- Tasks involving visual stimuli and attention: all CPT conditions and SDRT Cue
- N1/P2 complex widely identified in tasks involving visual stimuli and attention
- pERPs 5 and 6 combined correspond to the N1/P2 complex
- Significant loadings on pERPs 5 and 6 for each of the tasks locked to a visual stimulus at electrode Cz



# Application to ADHD data: Rare event

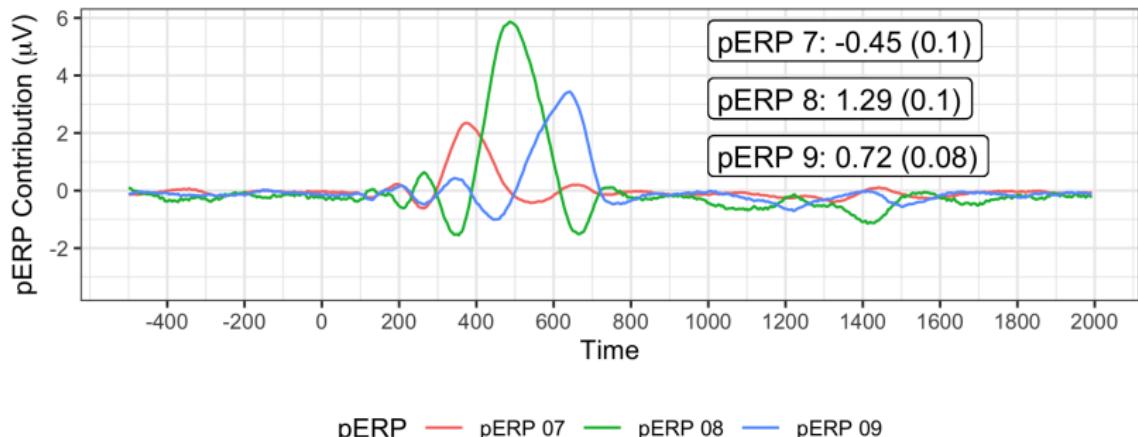
- CPT observed ERP X vs Not X at Cz



- Trials with an X are relatively rare → expected to produce a novelty signal typically associated with the P3
  - Expect activity at 300-500ms in contrast of Not X Incorrect and X Correct
- Contrast X Correct and Not X Incorrect since they both do not involve motor movement in the response

## Application to ADHD data: Rare event

- pERP contribution to the long positivity in 'CPT X Correct' task at Cz, with mean(SD)



- Contrast is explained by pERPs with peaks at different latencies
- All trial types with an 'X', regardless of motor response, show heavy loadings on these three pERPs → consistent with the expectation that pERPs 7, 8, and 9 relate to a novelty signal

## Application to ADHD data: APSD example

- APSD characterizes how loadings on a given pERP vary from one participant to the next
- ADHD has implications of the ability to maintain attention and working memory → expect more heterogeneity
- TD and ADHD-inattentive groups have similar levels of heterogeneity in loadings
- ADHD-combined (inattentive and hyperactive) has higher APSD values on every pERP, especially the first two and 10-15
- pERPs 1 and 2: reflect activity before the cue has disappeared from the screen
- pERPs 10-13: contribute to a late ongoing positivity in the waveform – perhaps related to maintenance of task-relevant attention or working memory

# Application to ADHD data: APSD example

Table 1: Maintenance Condition APSD (Cz)

pERP	Combined	Inattention	TD
pERP 01	<b>2.94</b>	1.50	1.44
pERP 02	<b>3.01</b>	1.68	1.69
pERP 03	1.91	1.41	1.44
pERP 04	1.07	0.94	0.90
pERP 05	1.15	0.90	0.87
pERP 06	1.17	0.94	0.90
pERP 07	1.29	1.02	0.86
pERP 08	1.34	0.94	0.93
pERP 09	1.56	0.79	0.80
pERP 10	<b>1.83</b>	0.88	0.74
pERP 11	<b>1.18</b>	0.70	0.68
pERP 12	<b>1.55</b>	0.65	0.56
pERP 13	<b>1.97</b>	0.91	1.07
pERP 14	<b>1.56</b>	1.25	1.01
pERP 15	<b>2.01</b>	1.42	1.38

## Simulations

# Data generation

- Data generation model

$$Y_{i,v,e}(t) = \sum_{p=1}^P k_{p,v,e} \phi_p^*(t) + \sum_{p=1}^P \xi_{p,i,v,e} \phi_p^*(t) \\ + \sum_{\ell=1}^L \alpha_{\ell,i,v,e} \psi_{\ell}(t) + \zeta_{i,v,e}(t)$$

- Total number of:

- tasks  $V = 9$
- true pERPs  $\phi_p^*(t)$   $P = 5$
- time points  $T = 500$
- Fourier bases  $\psi_{\ell}(t)$   $L = 7$
- electrodes per subject  $E = 40$

- Coefficients distributed as

- $k_{p,v,e} \sim N(0, 0.25)$
- $\xi_{p,i,v,e} \sim MN_{V,E}(0, \Sigma_{p,v}, \Sigma_{p,e})$
- $\alpha_{p,i,v,e} \sim MN_{V,E}(0, \Sigma_{\ell,v}, \Sigma_{\ell,e})$
- $\zeta_{i,v,e}(t) \sim N(0, \sigma_{\text{error}}^2)$

# Data generation

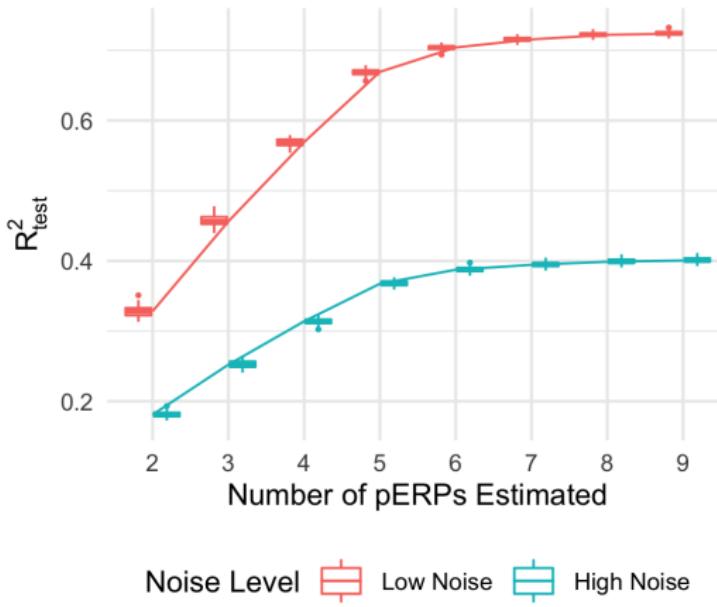
- Data generation model

$$Y_{i,v,e}(t) = \sum_{p=1}^P k_{p,v,e} \phi_p^*(t) + \sum_{p=1}^P \xi_{p,i,v,e} \phi_p^*(t)$$
$$+ \sum_{\ell=1}^L \alpha_{\ell,i,v,e} \psi_{\ell}(t) + \zeta_{i,v,e}(t)$$

- First term: ERPs at each task and electrode are a weighted average of the true pERPs  $\phi_p^*(t)$
- Second term: subject-specific deviations from task- and electrode-specific signal
- Third term: Fourier bases are used as noise structured in time
- Fourth term: random measurement error

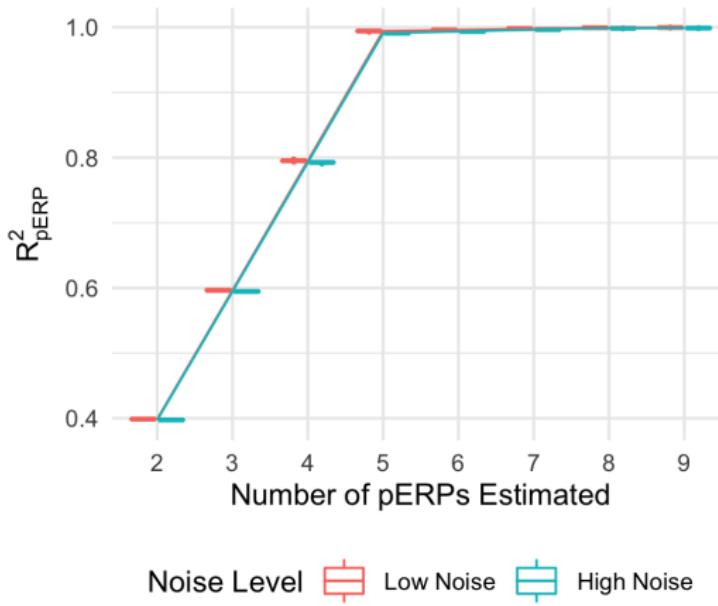
# Simulation example

- The elbow of the  $R^2_{\text{test}}$  is at 5, the true number of pERPs



# Simulation example

- $R^2_{\text{pERP}}$ : proportion of variation in the true pERPs explained by the estimated pERPs
- The elbow of the  $R^2_{\text{pERP}}$  is also at 5, the true number of pERPs



# Matching pERPs

- Regression coefficients from regressing the true pERPs on the estimated pERPs
- Each true pERP simulated was fitted by a combination of the estimated pERPs

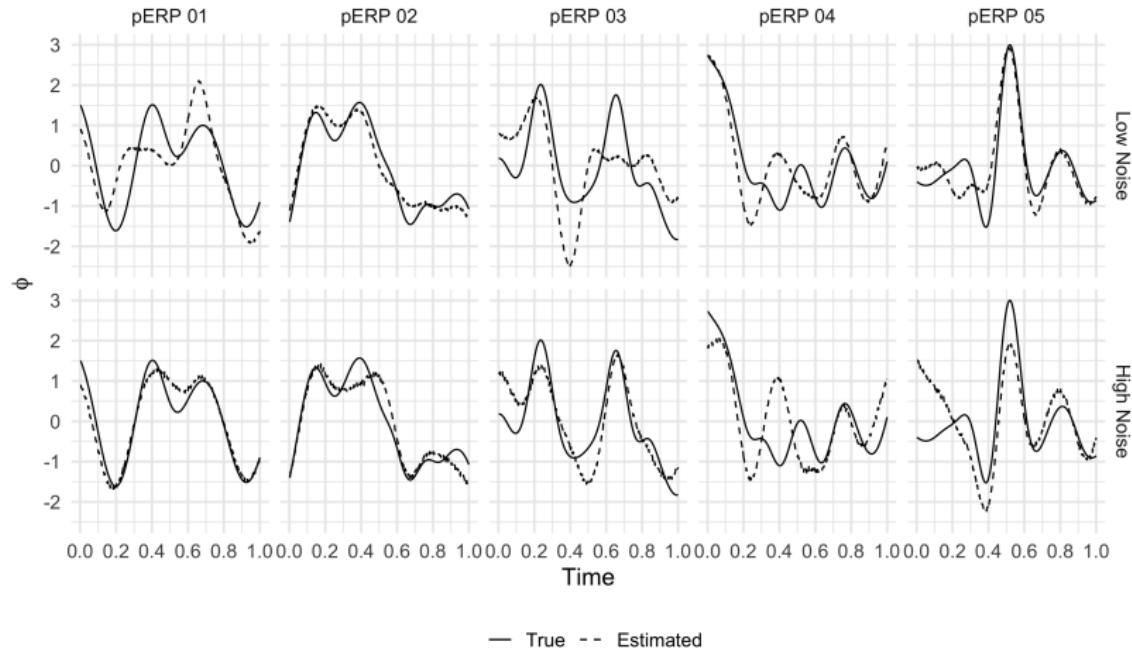
	pERP 1	pERP 2	pERP 3	pERP 4	pERP 5
pERP 5	0	0.94	0.15	0.04	0.28
pERP 4	0.11	-0.04	0.87	0.32	-0.34
pERP 3	0.05	0.25	-0.04	-0.65	-0.7
pERP 2	0.96	-0.06	0.02	-0.17	0.21
pERP 1	-0.25	-0.19	0.47	-0.66	0.5
True					

Figure 1: High noise

	pERP 1	pERP 2	pERP 3	pERP 4	pERP 5
pERP 5	0.75	-0.13	0.63	-0.15	0.05
pERP 4	-0.24	-0.16	0.29	-0.11	-0.9
pERP 3	0.47	0.17	-0.62	-0.52	-0.29
pERP 2	0.01	-0.96	-0.24	-0.08	0.11
pERP 1	0.4	-0.03	-0.27	0.83	-0.29
True					

Figure 2: Low noise

# Estimated pERPs



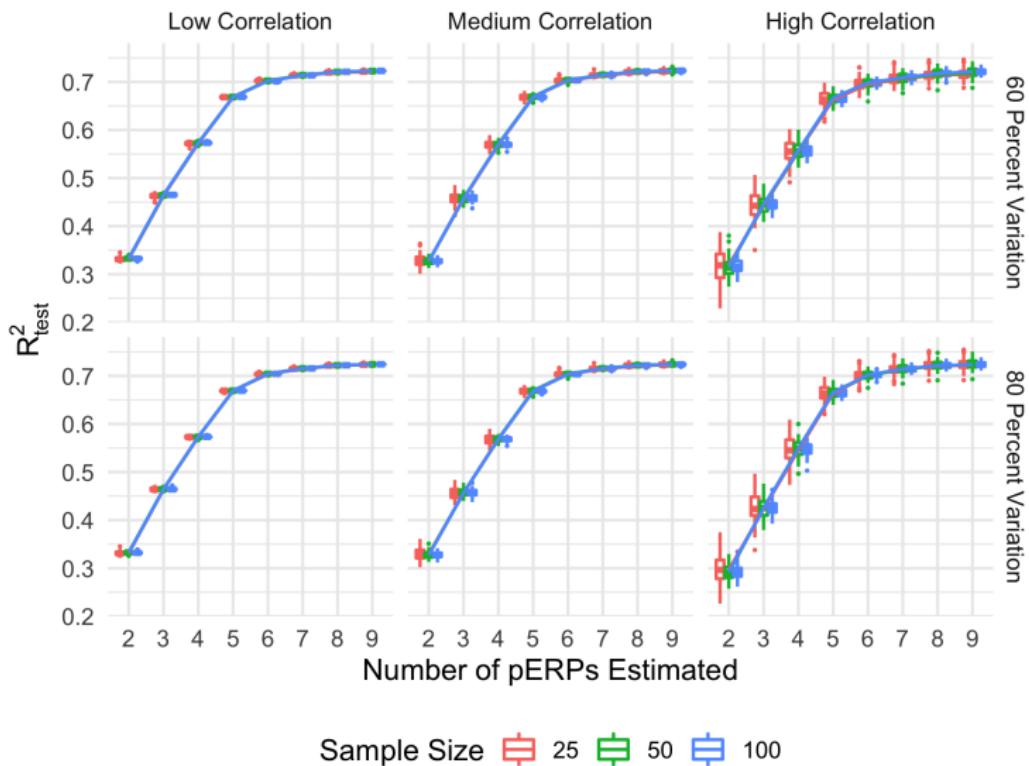
# Simulation Study

- Simulation cases:
  - Sample size  $N = 25, 50$  and  $100$
  - Correlation among electrodes and tasks  $\rho = 0.1, 0.5$ , and  $0.9$  (used in generation of  $\xi$  and  $\alpha$ )
  - Percent of variation used for retaining components in the PCA steps
  - Both the low and high noise cases,  $\text{SNR} = 1$  and  $0.6$
- Effects were assessed on  $R_{\text{pERP}}^2$  and  $R_{\text{test}}^2$
- Parameters varied did not affect the algorithm's ability to recover the true pERPs where  $R_{\text{pERP}}^2$  displayed the same pattern as the previous slide in all simulation cases

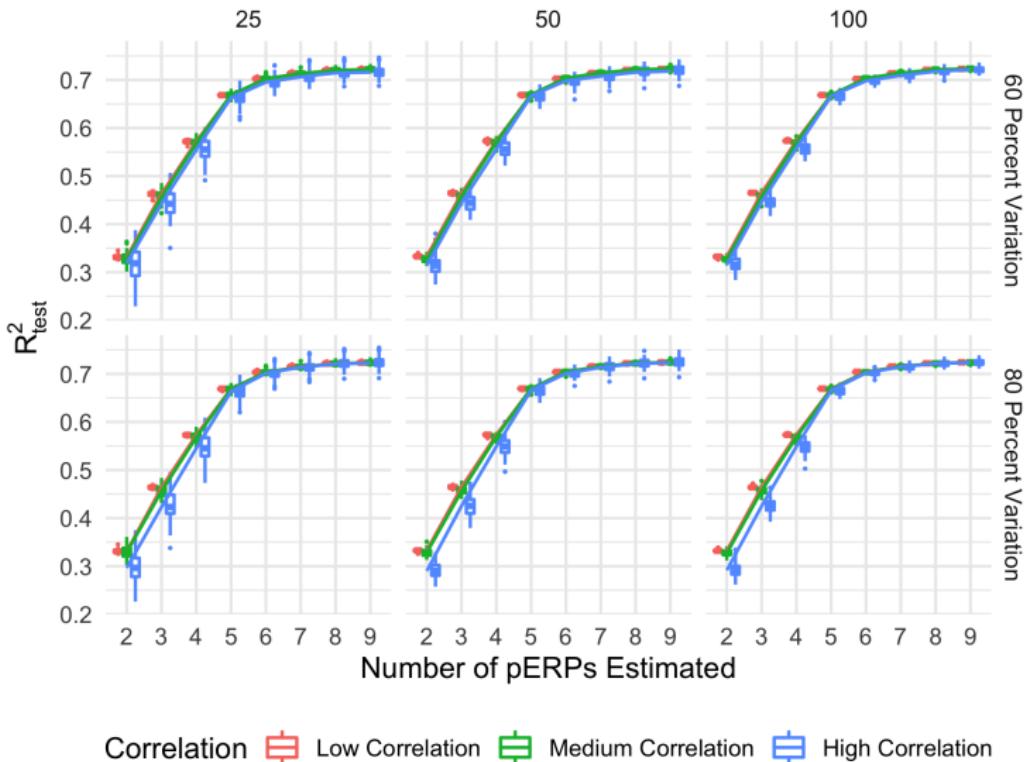
## Effects of sample size, correlation, and percent variation

- $R_{\text{test}}^2$  steady across sample sizes, but variability in prediction accuracy decreases with increasing sample size
- Higher correlations across electrodes and tasks correspond to smaller effective total number of electrodes → prediction accuracy gets worse (with greater variability)
- Percent of variation does not appear to affect the  $R_{\text{test}}^2$  except for the high correlation case
- High correlation among electrodes and tasks → retaining more variation in PCA corresponds to retaining more noise, leading to worse prediction accuracy

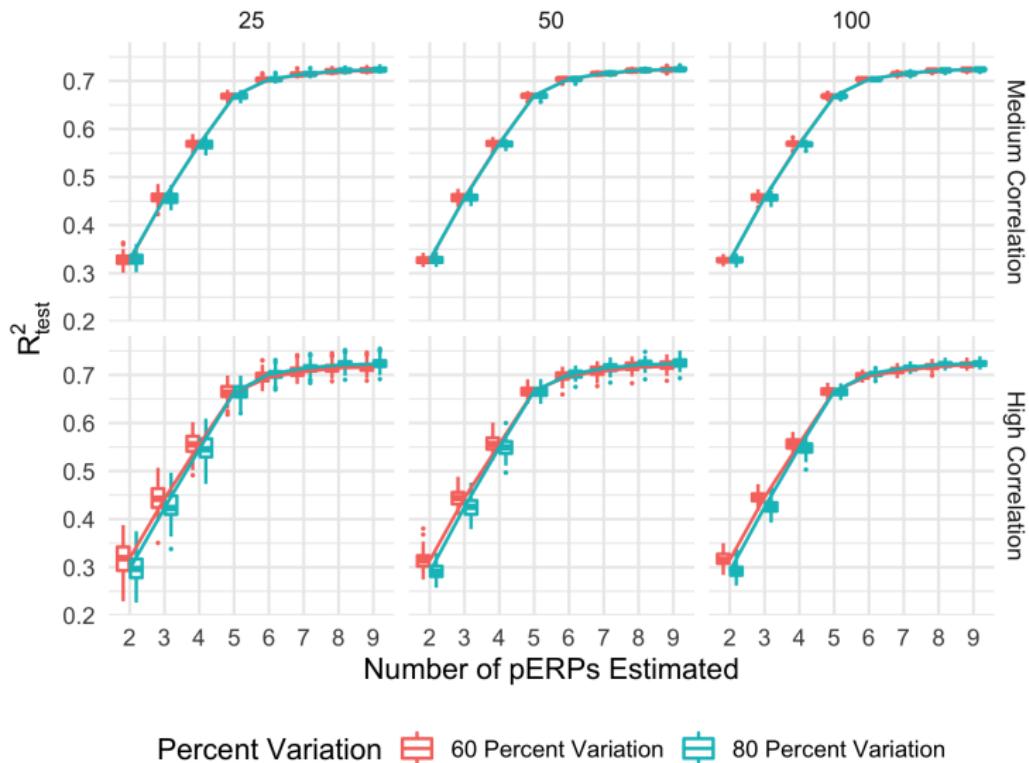
# Effects of sample size



# Effects of correlation



# Effects of percent variation

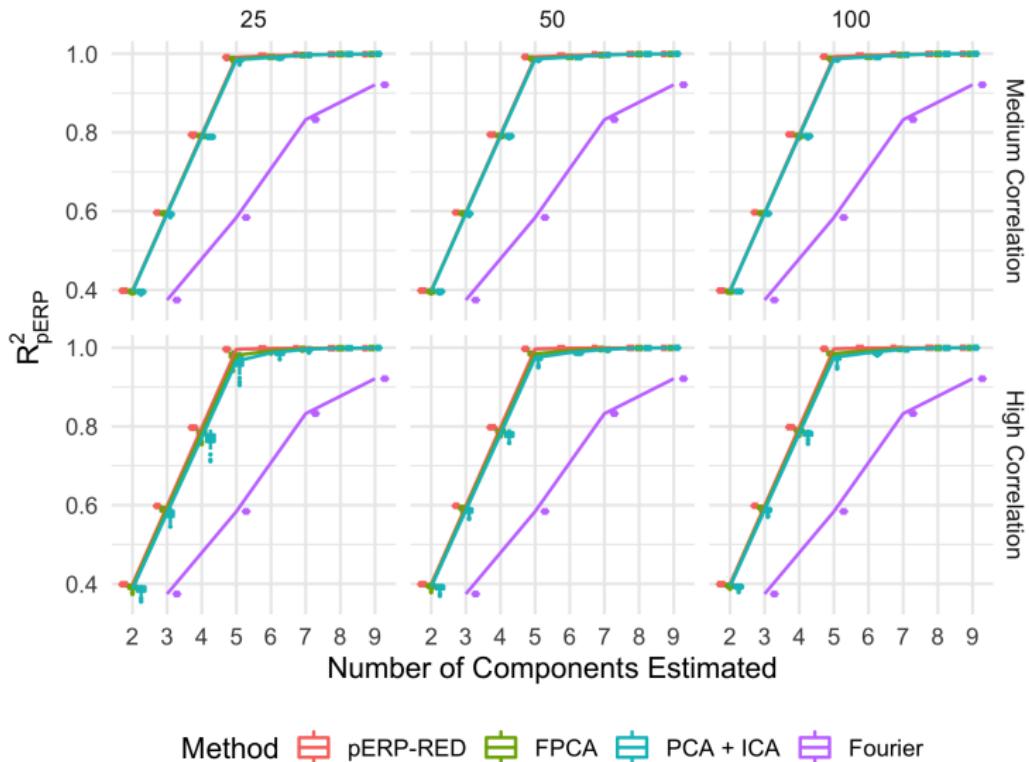


## Methods Comparisons

- Compared to: Fourier, functional principal component analysis (FPCA), and a “single-PCA” version of pERP-RED
- Fourier: bases are fixed rather than being data-driven
  - Many more components required to achieve the same predictive accuracy
- FPCA: seeks the greatest variation in the data with the fewest components, whereas pERP-RED uses ICA to extract maximally unmixed underlying signals
  - Very similar predictive accuracy with FPCA having slightly lower  $R^2_{\text{pERP}}$  and pERP-RED having slightly lower  $R^2_{\text{test}}$

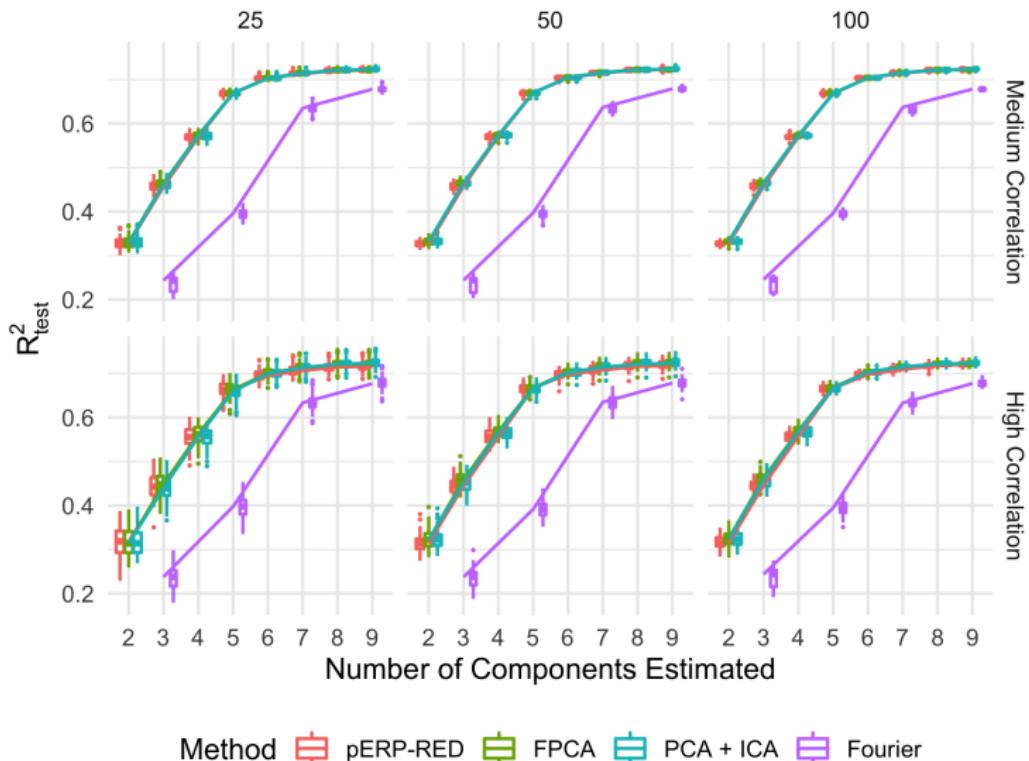
- “Single-PCA”: single PCA on the matrix formed using all of the subjects and electrodes in the columns
  - Similar performance in this setting on values of  $R_{\text{pERP}}^2$  and  $R_{\text{test}}^2$
  - In practice: single-PCA is limited by size of data, i.e.  
$$N \times \sum_{i=1}^N E_i \leq V \times T$$
  - Assumes heterogeneity in features across subjects, where pERP-RED does not
  - pERP-RED allows the user to control the amount of variation used in each PCA step separately

# Comparisons to other bases: pERP prediction



Method pERP-RED FPCA PCA + ICA Fourier

# Comparisons to other bases: Record prediction



## Concluding Remarks

- Developed a method for estimating an underlying set of components
- Provided tools for analyzing ERPs in terms of these components
- Developed the pERPred R package and designed (beautiful) Shiny applications to display results

- Extract the weight on a given pERP at the trial-wise level → predict or relate to trial-wise behavior
  - Model longitudinal trends over trials: (Scheffler et al., 2017), (Scheffler et al., 2019), (Fiecas et al., 2016), (Ombao et al., 2018)
- A functional data analysis approach
  - Recent work modeling multivariate hierarchical functional data have not considered data structures across multiple experiments, groups, subjects, electrodes, tasks and conditions: (Di et al., 2009), (Shou et al., 2015), (Happ et al., 2018), (Zhang et al., 2019)

## References

- C. Distefano, D. Senturk, S. S. Jeste, Erp evidence of semantic processing in children with asd, Developmental Cognitive Neuroscience 36 (2019) 100640.
- A. Lenartowicz, H. Truong, G. C. Salgari, R. M. Bilder, J. McGough, J. T. McCracken, S. K. Loo, Alpha modulation during working memory encoding predicts neurocognitive impairment in adhd, Journal of Child Psychology and Psychiatry.

# Thank you!

Slides available at [bit.ly/pERPred-talk](https://bit.ly/pERPred-talk)  
R package available at [bit.ly/pERPred](https://bit.ly/pERPred)