A simulation-based approach to statistical power with ERPs

Chia-Wen Lo & Jonathan R. Brennan

University of Michigan

chiawenl@umich.edu

Introduction

- Replicability in Event-related potential (ERP) studies have drawn increased attention recently [1]-[4].
- However, few studies examine how large is the effect size that can be reliably detected with standard ERP analyses (e.g. [1]).

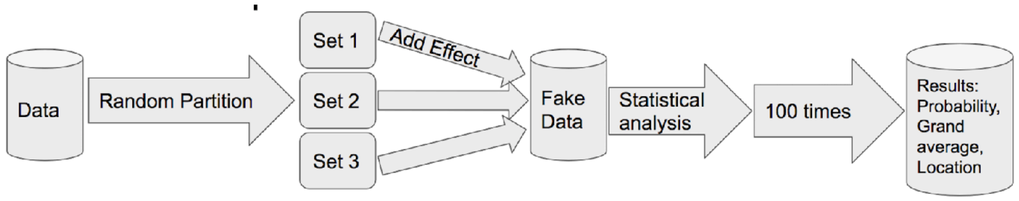
- Why replicability in ERP studies is difficult?
  - ERPs are a spatial-temporal matrix that reflect multiple parameters (amplitude over time, latency and duration for a given component, topographical distribution…). Thus, how to test that an effect "replicates" across different studies is quite difficult.
  - Though most literature reports the F-value or T-value along with their p-value, only 40% of papers report effect sizes, 56% reported mean values, and 47% reported some estimate of variance [5]. Rarely reporting such information impedes sample size calculation needed for conducting power analyses [6].
  - Existing toolboxes, like *Besa Simulator* or *Fieldtrip*, simulate ERPs from scratch by making assumptions about the source model and noise model which may not reflect actual data.

We provide a way to assess statistical power across a range of effect sizes and the number of participants, using both a P600 and a N400 as case studies.

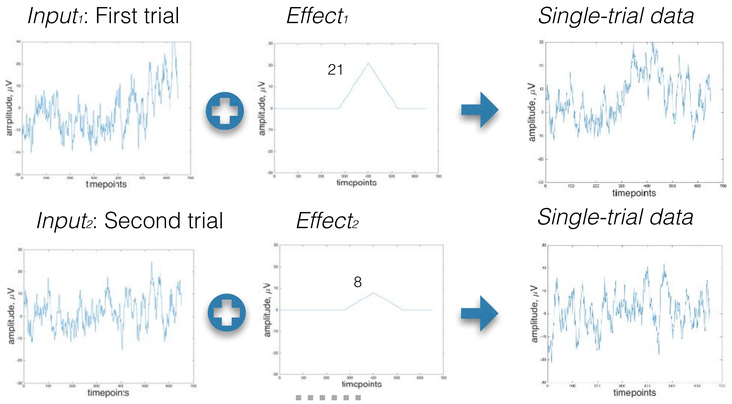
- We use actual raw single-trial data as the bases for our simulations, allowing for greater fidelity between simulated outcomes, and actual experimental outcomes.

Procedure

Pipeline

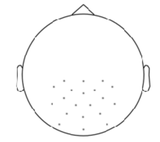


The distribution of **effect *E*** is parameterized in the following:
(i) Amplitude, (ii) Latency, (iii) Duration, and (iv) Topography.

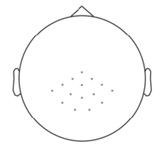


The simulation is conducted in **four stages**:
(i) Single-trial data from an experiment, already preprocessed, are randomly divided into partitions that reflect conditions for each participant.
(ii) A stochastic effect *E* is drawn from a Gaussian distribution and added to each trial in one partition.
(iii) The data are averaged and a group analysis is conducted as it would be for a typical experiment.
(iv) steps (i-iii) are repeated, yielding a distribution of statistical outcomes where the "true" effect *E* is known.

Posterior sensors for the P600:



Central-posterior sensors for the N400:



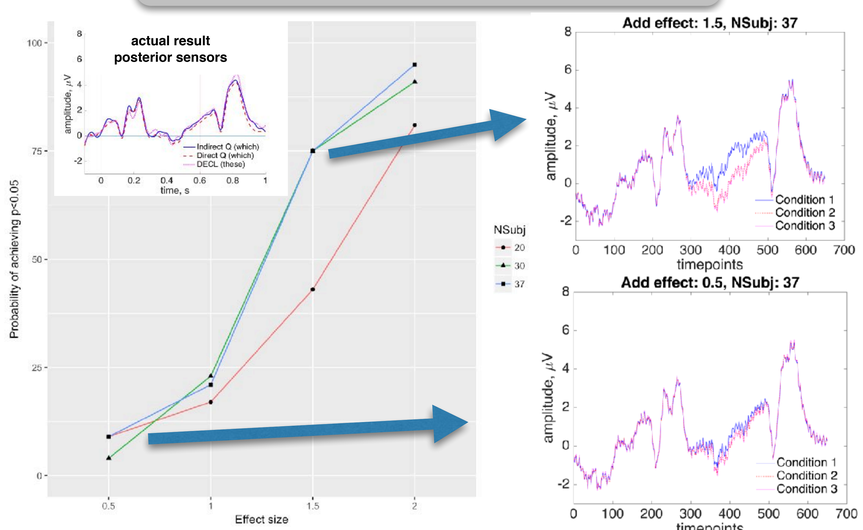
EEG Methods

EEG data were recorded at 500 Hz from 61 active electrodes. Epochs around word onset were re-referenced to linked-mastoids, cleaned of artifacts with visual inspection and ICA, band-pass filtered from 0.1–30 Hz, and baseline-corrected. A non-parametric statistical analysis was conducted across all electrodes time-locked to the target word for each experiment.

Simulation Results

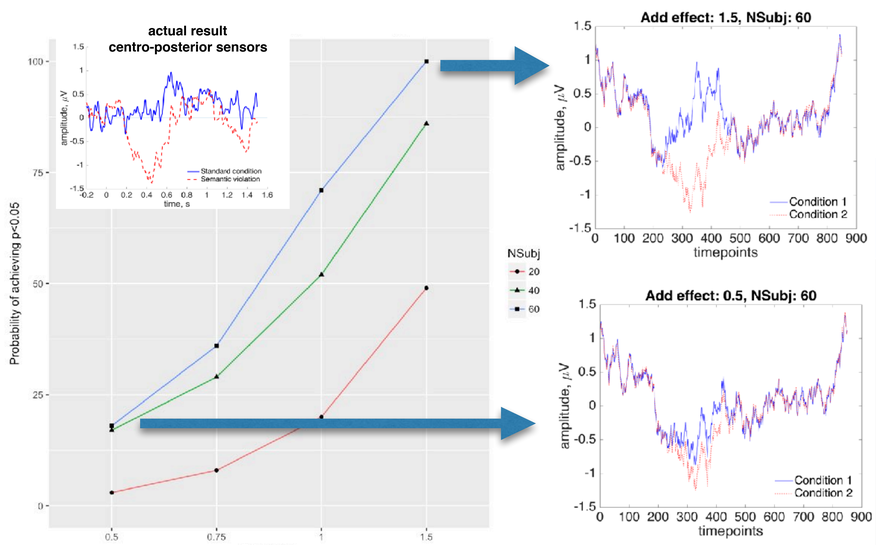
Case Study 1: P600

This Mandarin Chinese experiment tests for a P600 at the *wh*-word in the indirect questions, relative to the declarative sentences. N=37 subjects **read 64** sentences per condition. The experimental results show that no P600 effect was elicited at the *wh*-word, compared to the demonstrative word.



Case Study 2: N400

These data come from a standard N400 semantic violation protocol. N=60 subjects **listened to 40** sentences per condition. Semantic violations appeared sentence-medially.



Conclusions

- We present a simulation-based approach to quantifying statistical power for ERP studies.
- The current approach adds an effect equally to the relevant electrodes with the latency and duration fixed — we are also developing more flexible methods, e.g. computing effect by assigning different weights for duration and each electrode.
- The approach provides a way to test whether the effect is indeed absent by estimating whether different effect sizes suggested in the literature can be detectable in the data.
- The study also provides a way to estimate the number of participants and the effect sizes for the future relevant studies.

- The Matlab code will be available soon: https://gitlab.com/chiawenl/EffectSizeDetectSimulation

References [1] Boudewyn, M. et al. (2017). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology 55(6).* [2] Cohen, M.X. (2017). Rigor and replication in time-frequency analyses of cognitive electrophysiology data. *International Journal of Psychophysiology 111*: 80-87. [3] Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology 54*: 146-157. [4] Thigpen, N. N. et al. (2016). Assessing the internal consistency of the event-related potential: An example analysis. *Psychophysiology 54*: 123-138. [5] Larson, M. J., & Carbine, K. A. (2017). Sample size calculations in human electrophysiology (EEG and ERP) studies: A systematic review and recommendations for increased rigor. *International Journal of Psychophysiology 111*: 33-41. [6] Guo, Q. et al. (2014). A systematic review of the reporting of sample size calculations and corresponding data components in observational functional magnetic resonance imaging studies. *NeuroImage 86*: 172-181.