ERP DATA ANALYSIS, TRADITIONAL, PCA, AND FDR

Jia Wu PhD, Michael J. Crowley PhD, Linda C. Mayes MD

Yale Child Study Center New Haven, CT

Introduction

EEG has rapidly changed from using only a few channels to hundreds in recent years. To be conservative, many researcher still focus their analysis on a few well characterized channels (Fz, Cz, Pz etc), without taking advantage of the multi-channel net.

New research also tends to analyze time periods that have been used by previous literature, even though it might not be the most appropriate to the data set. Recognizing these issues, investigators attempt to address windowing and localization problems with new methods, all of which introduce new problems to consider.

In this study, we compared four different data analysis methods on the same dataset. Results from all the methods are compared.

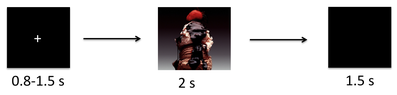
A. Traditional approaches limit in predefined time windows and channel clusters.
B. Pre-defined channel cluster and temporal PCA (Molfese, et al, 2006).
C. Spatial Temporal PCA (Dien, 2010) blindly cluster data based on the variance distribution.
D. False Discovery Rate (FDR) Benjamini & Hochberg (1995) and the Benjamini & Yekutieli (2001) utilized multiple comparisons in all the spatial/temporal dimensions and uses a statistical way to correct it.

Methods

Subjects: n=69, male=40, age range: 10-17 years old, mean = 13.57, sd = 2.15. Twelve subjects did not produce sufficient artifact free trials and were excluded from the analysis.

Task:

45 food and nonfood images mixed



EEG: recorded from 128 channel HydroCel (EGI) from DC Amp 300. Sampling rate 250 Hz, 0.1-100 Hz. Data was preprocess through 0.1 Hz high pass, 30Hz low pass filter, segmentation to 100 ms pre and 900 ms post the stimulus onset. Bad eye channels were detected manually and replaced. Then bad segments and channels of the whole dataset were detected by Artifact detection. All trials with any eye blink or eye movement (>150 μV) or with more than 10 bad channels (>200 μV) were rejected. Subjects with equal or more than 15 trials were retained for future analysis.

Method A: Following previous literature of the same paradigm that looked for the condition difference in P3 (300 – 400 ms) and LPP (400 – 800 ms) in the left and right anterior, central and posterior regions (Nijs, Franken, Muris 2008).

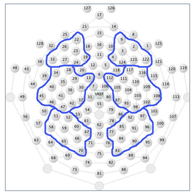
Method B: Average data across defined clusters (left and right Frontal, Central, Parietal, Occipital and Temporal lobes), then apply Factor Analysis (PCA followed by Varimax Rotation) on the temporal dimension to determine the time range. Finally average the data on the factor based time ranges.

Method C: Apply PCA (Promax rotation) on the temporal dimension of all the channels first, then apply PCA (Informax rotation) on the spatial dimension to generate a factor based cluster for each temporal PCA component.

Method D: Calculate the condition difference at each data-point on each channel, then use FDR to find the significant ones. Finally use PCA to locate the common window of all the significant channels.

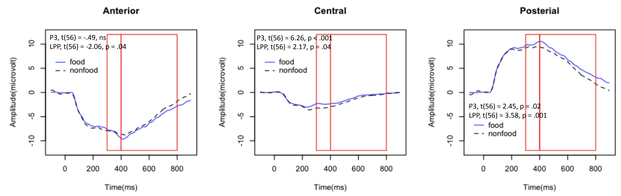
Results

A. Traditional 3x2 cluster, P3 and LPP:

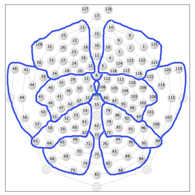


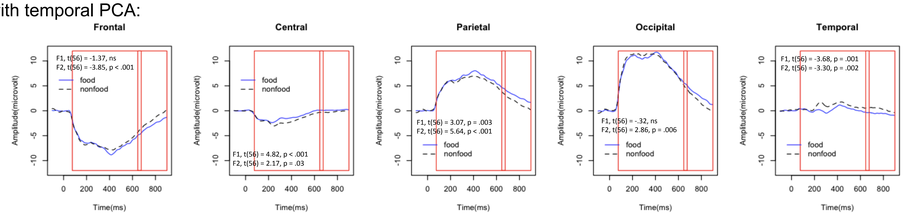
P3: ANOVA on region x hemisphere x condition:
Region, F(2,112) = 208.56, p<.001. Hemisphere, F(1, 56) = .01, ns. Condition, F(1, 56) = 50.75, p<.001. Region x hemisphere, F(2,112) = 12.67, p<.001. Region x condition, F(2,112) = 3.83, p=.048. Hemisphere x condition, F(1, 56) = .17, ns. Region x hemisphere x condition, F(2, 112) = 2.82, ns.

LPP: ANOVA on region x hemisphere x condition:
Region, F(2,112) = 159.29, p<.001. Hemisphere, F(1, 56) = 2.02, ns. Condition, F(1, 56) = 8.77, p=.004. Region x hemisphere, F (2,112) = 4.72, p=.02. Region x condition, F(2,112) = 7.49, p=.005. Hemisphere x condition, F(1, 56) = 7.13, p=.010. Region x hemisphere x condition, F(4, 224) =.55, ns.



B. Pre-defined cluster with temporal PCA:





Temporal F1 (76-676 ms): ANOVA on region x hemisphere x condition. Region, F(4,224) = 179.33, p<.001. Hemisphere F(1, 56) = .02, ns. Condition, F(1, 56) = .01, ns. Region x hemisphere F(4,224) = 9.14, p<.001. Region x condition, F(4,224) = 6.42, p=.003. Hemisphere x condition, F(1, 56) = 3.75, ns. Region x hemisphere x condition, F(4, 224) = .53, ns.
Post Hoc of Region x condition interaction showed that Central, Parietal and Temporal had a significant condition difference.

Temporal F2 (644 – 900 ms): ANOVA on region x hemisphere x condition. Region, F(4,224) = 35.34, p<.001. Hemisphere, F (1,56) = 4.14, p<.05. Condition, F(1, 56) = 2.97, ns. Region x hemisphere, F(4, 224) = 1.28, ns. Region x condition, F (4,224) = 13.89, p<.001. Hemisphere x condition, F(1, 56) = 9.42, p=.003. Region x hemisphere x condition, F(4, 224) = 2.40, ns.
Post Hoc of Region x condition interaction showed that ALL regions had a significant condition difference.

C. Spatial Temporal PCA:

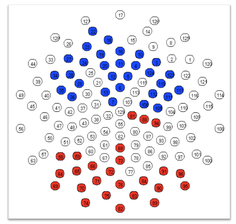


Fig. Spatial Factor 1 for temporal Factor 1 (68-660 ms)

Temporal F1 (68 – 660 ms):
ANOVA: cluster x condition
Cluster effect : F(1,56) = 201.95, p<.001.
Condition effect. F(1,56) = 4.26, p=.044.
However condition * cluster is not significant, F(1,56) = . 42, ns.

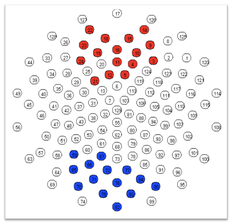
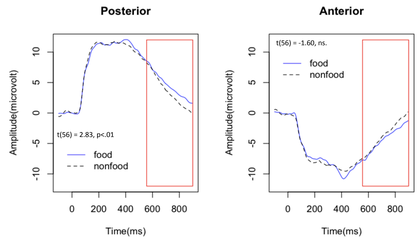


Fig. Spatial factor 1 for temporal Factor 2 (556-900 ms)

Temporal F2 (556 – 900 ms):
ANOVA: cluster x condition.
Cluster effect : F(1,56) = 59.26, p<.001.
Condition effect. F(1,56) = 1.33, ns.
Cluster x Condition interaction:
F(1,56) = 5.35, p=.02
Post Hoc on each cluster indicated that only the posterior site had a significant condition effect.



D. FDR

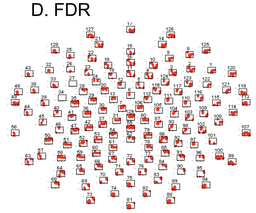


Fig. Without FDR, there were 10011 significant comparisons among 29025 total tests.

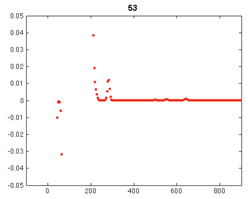


Fig. Without FDR, single channel p value.

FDR

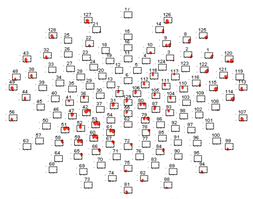


Fig. After FDR, there were 1678 significant comparisons among 29025 total tests.

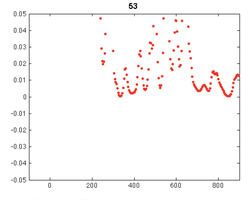
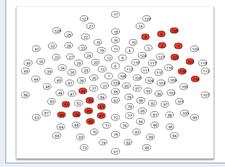
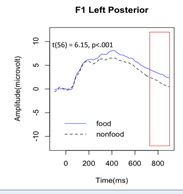
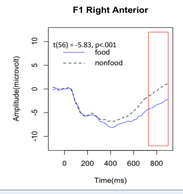


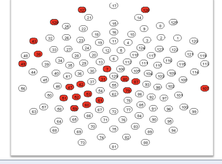
Fig. After FDR, single channel p value.



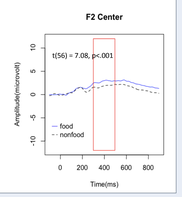
Temporal Factor 1: 728-900 ms. Two clusters. Cluster*Condition is significant. (F(1,56) = 47.07, p<.001). Post hoc tests in left posterior and right anterior were both significant.







Temporal Factor 2: 300-496 ms. The center cluster was chosen. T test of the condition effect was significant.



Conclusions/Discussions

1. All the methods were able to identify two components on the temporal dimension. The condition difference was significant in some of the cases for the early component (P3 related), and was always significant for the late component (LPP). The P3 and LPP effects were consistent with previous literature.

2. Method A (traditional P3 and LPP) arbitrarily defined the time range for P3 and LPP. From the plot, it could be observed that the time range for P3 did not always include the whole P3 peak. The LPP range also cut through the P3 peak in some of the situations. If the separation of P3 and LPP was important, defining the time based on previous literature was not a very effective approach.

3. Method B (pre-defined cluster with temporal PCA) was able to separate P3 and LPP correctly. It was a good way of dividing the whole head and studying each area. It also had the advantage of studying hemisphere effects. This approach missed the effect on the midline channels.

4. Method C (spatial-temporal PCA) identified similar components as method B. However because the spatial cluster was relatively large, the P3 effect was watered down and was not significant. The LPP effect was significant only in the posterior cluster.

5. Method D (FDR) was able to identify a narrow region for P3 in the posterior region, and a LPP region in both left posterior and right anterior regions. Comparing with the results from Method B, method D pointed out the most significant region on the scalp for both P3 and LPP.

6. The above discussions are based on a single, relatively large dataset. In other work with the medial frontal negativity, we have seen that method B has the drawback of merging unrelated channels, while method D may require large and/or reliable effects to detect differences. More comparisons of different methodologies needed to be done for different types of ERP components.

References

1. Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B (Methodological). 57(1), 289-300.
2. Benjamini, Y. & Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics. 29(4), 1165-1188.
3. Dien J, (2010), The ERP PCA Toolkit: an open source program for advanced statistical analysis of event-related potential data. Journal of neuroscience methods. Mar 15;187(1):138-45
4. Nijs IM, Franken IH, Muris P.(2008) Food cue-elicited brain potentials in obese and healthy-weight individuals. Eat Behav. 2008 Dec;9(4):462-70. Epub 2008 Aug 15.
5. Molfese, DL, Key, AF, Kelly, S, Cunningham, N, Terell, S, Ferguson, M, Molfese, VJ, Bonebright, T. (2006), Journal of learning Disabilities, Jul-Aug; 39(4): 352-63

Acknowledgements/Download

Support by Yale's Interdisciplinary Research Consortium on Stress, Self-Control and Addiction (Yale-IRC).

To download a color PDF version of the poster, please go to the website at the bottom of the poster, or scan the QR code on the right side:

http://medicine.yale.edu/labs/del/material/SPR_2011_Wu.pdf