

# ***Probability waves: pattern-based p-value correction in mass univariate analysis between two event-related potential waves.***

Dimitri Marques Abramov

Laboratory of Neurobiology and Clinical Neurophysiology

National Institute of Women, Children and Adolescents Health Fernandes Figueira, Oswaldo Cruz Foundation.

Address:

Avenida Rui Barbosa, 716. Bairro Flamengo, Rio de Janeiro, RJ

CEP 20021-140

dimitri.abramov@iff.fiocruz.br

Keywords: p-value correction, multiple comparisons, ERP statistics, univariate analysis, probability waves

## **Abstract**

**Background:** Methods for p-value correction are criticized for either increasing Type II error or improperly reducing Type I error. This problem is worse when dealing with hundreds or thousands of paired comparisons between waves or images which are performed point-to-point. This text considers patterns in probability vectors resulting from multiple point-to-point comparisons between two ERP waves (mass univariate analysis) to correct p-values. These patterns (probability waves) mirror ERP waveshapes and might be indicators of consistency in statistical differences.

**New method:** In order to compute and analyze these patterns, we convoluted the decimal logarithm of the probability vector ( $p'$ ) using a Gaussian vector with size compatible to the ERP periods observed. For verify consistency of this method, we also calculated mean amplitudes of late ERPs from Pz (P300 wave) and O1 electrodes in two samples, respectively of typical and ADHD subjects.

**Results:** the present method reduces the range of  $p'$ -values that did not show covariance with neighbors (that is, that are likely random differences, type I errors), while preserving the amplitude of probability waves, in accordance to difference between respective mean amplitudes.

36 **Comparison with existing methods:** the positive-FDR resulted in a different  
37 profile of corrected p-values, which is not consistent with expected results or  
38 differences between mean amplitudes of the analyzed ERPs.

39 **Conclusion:** the present new method seems to be biological and statistically  
40 more suitable to correct p-values in mass univariate analysis of ERP waves.

41

## 42 **Introduction**

43 When we analyze event-related potentials (ERP), we primarily focus on  
44 latencies and amplitudes of the arbitrarily determined elements in these  
45 waves. We generally calculate the maximum and mean amplitudes (mean of  
46 amplitudes within an interval) in these elements, and statistically infer the  
47 difference in these parameters between two samples of that wave. Wave  
48 latency is determined based on its maximum amplitude. These elements are  
49 the components (or waves) of a set of signals that are systematically  
50 observed in a population of individuals under the same experimental  
51 conditions. For instance, the P100 wave obtained from human brain activity  
52 under visual stimulation of a reverse pattern [1]. However, the definition and  
53 delimitation of this ERP is historically arbitrary "to the naked eye". Moreover,  
54 traditional p-correction analyses have lost several pieces of information  
55 regarding the identity of these waves, e.g., differences between periods and  
56 phases of these ERPs.

57 Mass univariate analysis (MUA) brings a new perspective to assess ERP  
58 behavior, as it consists of describing the differences between two waves that  
59 are explored when they are compared point to point, and we plot the statistical  
60 differences, using a method we call here raster pairwise comparison (RPC),  
61 which returns a raster diagram of p-values as a function of time [2]. See figure  
62 1. Hence, by using RPC we can observe the difference between waves over  
63 time, combining latencies and amplitudes in one single measurement [3].

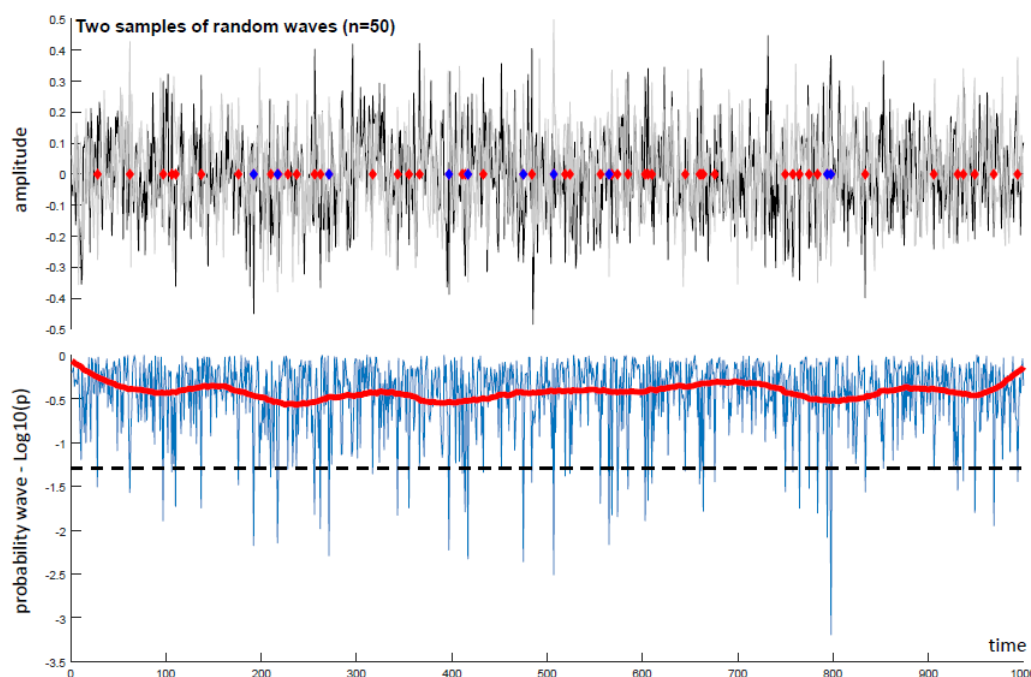
64 These pairwise comparisons can be tested with the most suitable statistical  
65 method according to sample features. In MUA, hundreds, even thousands of  
66 comparisons are performed according to the temporal extensions of waves.  
67 These comparisons are explanatory, i.e., we broadly seek for differences.  
68 This brings us to the issue of multiple comparisons, which substantially  
69 increase type I errors [4]. Historically, the concern with spurious differences  
70 has led to the development of methods to correct the probability of having  
71 differences due to the number of comparisons. The most traditional method is  
72 Bonferroni correction. However, as other methods of the same class, it is  
73 quite conservative [4].

74 Benjamini & Hockberg developed a method of False Discovery Rate (FDR),  
75 which is less conservative and is indicated, along with its variations, for

multiple comparisons of MUA scale [5, 6]. After, other variants of FDR method were developed, which concern interdependence among comparisons [7, 8]. These correction methods are used to compare, e.g., genomes and resonance imaging (pixel to pixel). We used this method in a previous study for a point-to-point comparison of ERP waves (RPC).

Although FDR methods presume that pairwise compared data vectors are correlated to each other in terms of covariance (which would explain the fact that they correct p-values with a lower degree of rigor), these methods do not consider covariance of sampling vectors when calculating p-value correction.

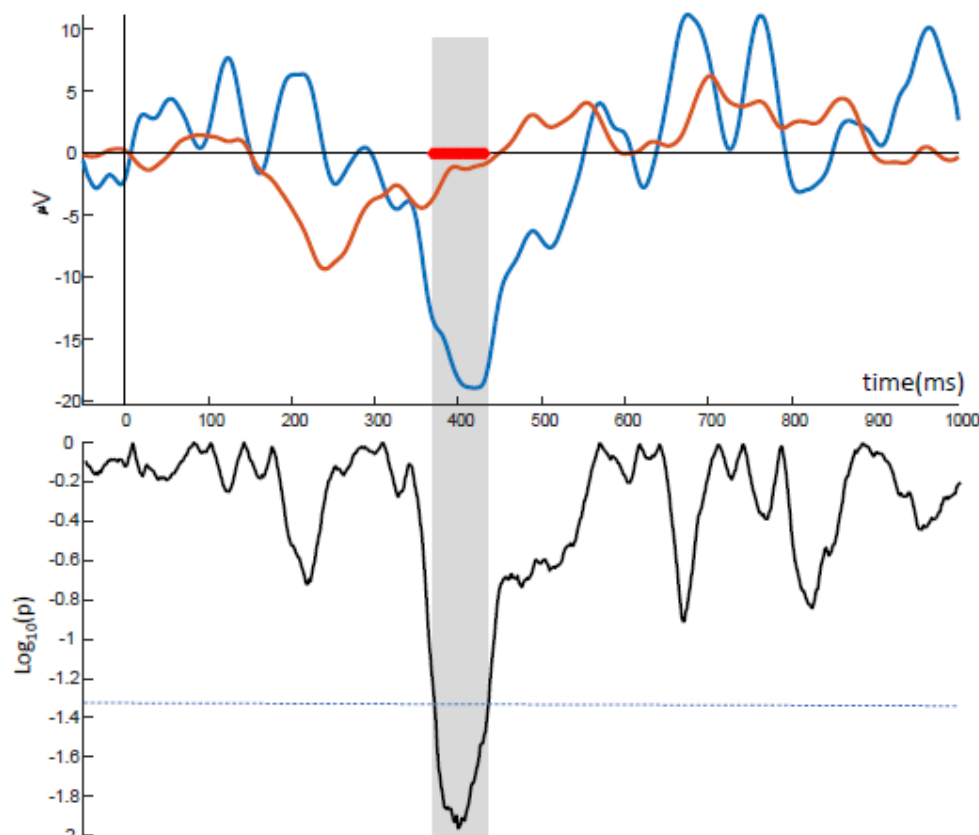
We propose here a new paradigm for the correction of multiple comparisons in test sets where the variables tested have a natural correlation to each other, which can be considered *a priori*. P-values form a probability vector with behavioral patterns that might indicate that differences are statistically consistent in wave regions with  $p \leq \alpha$ .



**Figure 1. Mass univariate analysis with and without p-correction in a random scenario.** Top: raster diagram of multiple pairwise comparisons (t-tests) between two samples with 50 random waves each (means in black and grey), and 1000 time points. Each blue diamond represents  $p \leq 0.01$ , while the red ones represent  $p \leq 0.05$  regarding the null hypothesis for each pair of point vectors. We found 57 significant differences ( $p < 0.05$ , ~ 5% of all points). Bottom: non-corrected probability vector (blue) and corrected one (red) by the present method, observing that all significances were rejected.

# 101 **Theory and Method Application.**

102 A probability vector might have a visibly stochastic profile (figure 2), in which  
 103 the distribution of statistically significant p-values does not follow any pattern.  
 104 In order to observe this behavior, we derived the original p-value vector into a  
 105  $p'$  vector, where  $p' = \log_{10}(p)$ . This vector, as a whole, suggests that these  
 106 significant p-values might be erratic (i.e., type I errors).



107  
 108 **Figure 2. Probability waves.** Vector of probability of rejecting the null hypothesis  
 109 (bottom), from the raster pairwise comparison between two ERPs (top), showing  
 110 collective behavior patterns of  $p'$ -values, organizing “probability waves” with profiles  
 111 similar to the ERP. Shaded areas are the ones where the null hypotheses were  
 112 rejected by  $\alpha = 0.05$  (red diamonds).  $p' = \log_{10}(p)$ .

113  
 114 However, another probability vector might show patterns of order that denote  
 115 covariance between sampling data of neighbors (figure 3). This pattern  
 116 observed is the gradual and massive evolution of  $p'$ -values forming a  
 117 “probability wave” equivalent to that of the analyzed ERP (figure 3, dashed  
 118 window).

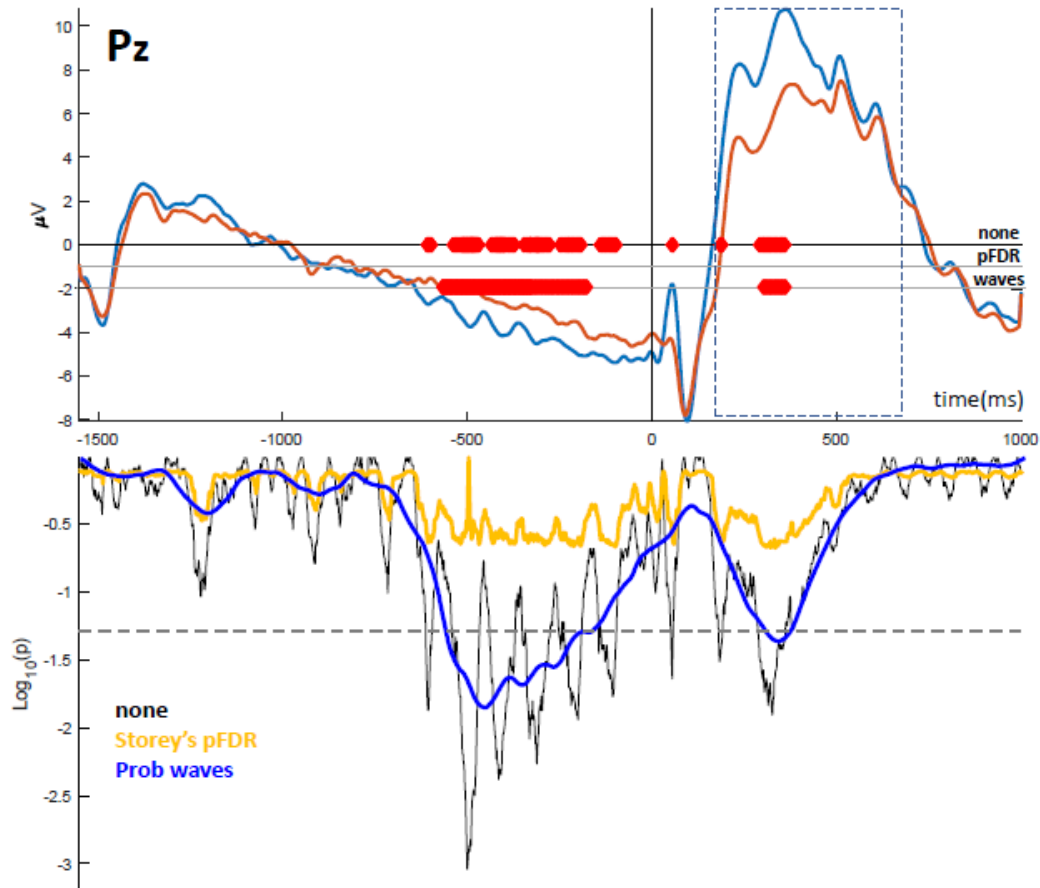
119 We are not analyzing these behavioral patterns with unaided eye. One way of  
 120 isolating such patterns is using mathematical convolution of  $p'$ -value in the  $t_o-t$   
 121 interval, using a vector of n-values, according to the equation:

122  $p''(t_o-t) = p'(t_o-t) * \mathbf{n}$

123 where  $t_o = [1, 2, \dots, t_{max}-t]$ ,  $t_{max}$  is the total size of the wave under analysis (in  
124 time points) and  $\mathbf{n}$  is a value vector that sets one regular curve with size  $t$ .

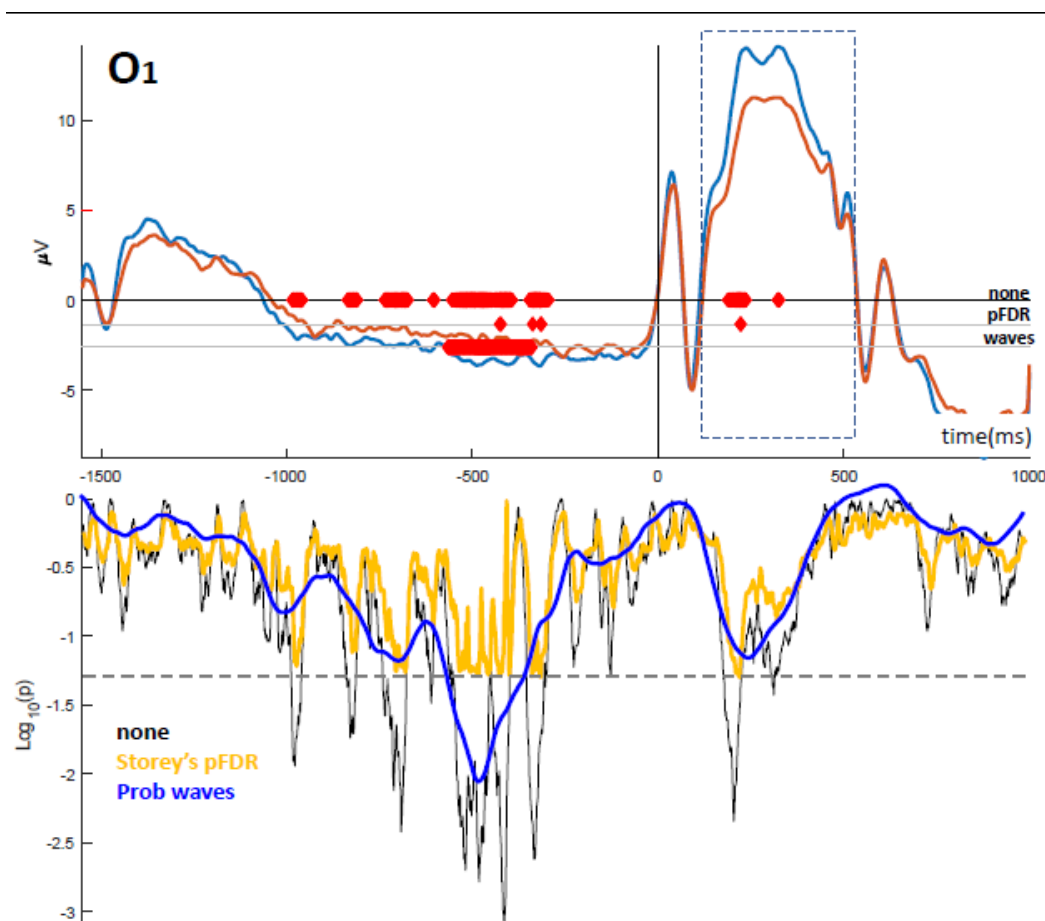
125  $\mathbf{n} = \exp(-\mathbf{x}^2/2)$

126 where  $\mathbf{x}$  is the vector  $[x_1, x_2, \dots, x_t]$ , where  $x_1 = -\sqrt{2}$ ,  $x_t = \sqrt{2}$ , and  $t$  is the number  
127 of points in the wave compatible with the period of ERPs to be compared. We  
128 used here 60 points (equal to 100ms, at a 600Hz sampling rate). In the  
129 convolution algorithm,  $\mathbf{n}$  slides on the  $\mathbf{p}'$  vector, thus creating a smoothing  
130 wave, where erratic values (with periods much lower than  $t$ ) tend to be  
131 suppressed. We studied the effect of the convolution of  $\mathbf{p}'$  by  $\mathbf{n}$  ( $t = 60$  bins ~  
132 100ms) on the probability waveform resulting from multiple U-tests (Mann-  
133 Whitney) between two ERP waves (case and control), from the O1 and Pz  
134 electrodes (10-20 montage), obtained in an experiment that evaluated  
135 neurophysiological correlates of behavior in the Attention Network Test of  
136 both typical ( $n = 20$ ) and ADHD children ( $n = 19$ ) [3]. We show the  
137 uncorrected probability waves (figure 3, bottom of Pz and O1 panels, in black)  
138 and those corrected by the proposed method (blue). We also used the Storey  
139 method for p-correction (positive False Discovery Rate, p-FDR, orange) [9],  
140 which is less conservative than the Benjamini-Hockberg [6] and Benjamini-  
141 Yekutieli [7,8] methods.



142

143



**Figure 3. Methods for p-correction in UVA.** Top: comparing waves from Pz electrode between groups (control in blue, cases in red), correlated to Attention Network Test behavior, performed by typical and ADHD youths (see text). P3 wave is inside the dashed window. The non-corrected probability wave (bottom, black) shows several rejected null hypotheses (red diamonds:  $p < 0.05$ ). P-correction was performed by positive FDR ("pFDR", orange) and the proposed method ("wave", blue). The gray dashed line corresponds to the statistical boundary ( $p < 0.05$ ). Bottom: for waves from O1 electrode.

To convolve over the probability wave, we used a normal curve with a period of  $t = 100\text{ms}$ , with order of magnitude of the expected waves and that is thus able to isolate patterns with periods equal to or greater than 100ms. To estimate the reliability of p-correction methods, we compared the mean amplitude of the cognitive target-related potentials (dashed line) on the Pz electrode, which is the P300 wave [10], and on the O1 electrode. We tested the null hypothesis between those mean amplitudes using the Mann-Whitney U Test.

164 Observing the effect of the proposed method on a stochastic scenario (figure  
165 1), all statistical differences were rejected, as they should be. After  
166 convolution, the resulting probability wave shows very low amplitudes (figure  
167 1, bottom, red wave).

168 In the real-life scenario, the mean amplitudes of the target-related potential  
169 from Pz were significantly different between groups (control:  $6.70 \pm 3.16 \mu\text{V}$ ;  
170 cases:  $4.18 \pm 4.24 \mu\text{V}$ ;  $p = 0.028$ ). However, significance was not observed in  
171 the corresponding differences from O1 electrode (control:  $9.66 \pm 2.52 \mu\text{V}$ ;  
172 cases:  $7.82 \pm 3.17 \mu\text{V}$ ;  $p = 0.053$ ). As observed in figure 3, the p-correction  
173 methods resulted in different  $\mathbf{p}'$ -vectors. Here, the proposed method rejected  
174 the null hypotheses, thus behaving consistently with the differences found  
175 between the mean amplitudes.

176

## 177 Discussion

178 The P3 wave is an attention-related cognitive neurofunctional complex that  
179 manifests on parietal site, and which appears around 300 milliseconds after  
180 the corresponding event [10]. Thus, we would expect some effect of the test  
181 on the P3 wave only on the Pz electrode. The proposed method corrected the  
182 p-values more consistently with the expected behavior of the P3 wave than  
183 the pFDR method. In the universe of event-related potentials, the method of  
184 probability wave using convolution of compatible magnitude to the biologically  
185 expected one was more reliable.

186 We presume that in Nature, collateral points of a biological wave are  
187 correlated to each other, as they derive from deterministic processes,  
188 although they have a chaotic nature. Thus, the behaviors of both equivalent  
189 waves, which are produced by the same source, follow the same causal  
190 mechanisms. The differences between these waves, statistically speaking,  
191 would also follow, by principle, a variation pattern that mirrors the profile of  
192 these waves.

193 Considering waves resulting from two different processes derived from the  
194 same causal mechanism (for instance, potential related to rare and frequent  
195 stimuli in an OddBall paradigm, resulting from different neural processes  
196 derived from the same neural mechanism [10]), theoretically, the chance of  
197 observing a false negative test result (type II error) is much lower than a false  
198 positive result (type I error). This is because different processes have a causal  
199 relationship with the same mechanisms (non-randomization).

200 Therefore, since a set of points is statistically different (rejecting the null  
201 hypothesis), because they correspond to the lowest values in a subset of  
202 points of the probability vector, which shows an organized behavioral pattern

(a probabilistic wave), these statistically determined differences might be considered to be true.

Hence, in a Mass Univariate Analysis between two ERP waves presumably derived from the same biological processes, values lower than  $\log_{10}(\alpha)$  of the probability pattern vector ( $p''$ ) do not correspond to type I errors.

## Acknowledgements

I would like to especially thank prof. Mario Fiorani Jr, whose computational function for convolution was used in this work. I would also like to thank the oswaldo cruz foundation for the opportunity to develop this work.

The author authors have no competing interests to declare.

## References:

- [1] Chiappa K. Evoked Potentials in Clinical Medicine. 3<sup>rd</sup> edition. Philadelphia: Lippincott Raven, 1997..
- [2] Groppe DM, Urbach TP, Kutas M. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology*. 2011 Dec;48(12):1711-25
- [3] Abramov DM, Cunha CQ, Galhanone PR, Alvim RJ, de Oliveira AM, Lazarev VV. Neurophysiological and behavioral correlates of alertness impairment and compensatory processes in ADHD evidenced by the Attention Network Test. *PLoS One*. 2019;14(7):e0219472.
- [4] Dawson B and Trapp RG. *Basic & Clinical Biostatistics*. 4<sup>th</sup> Ed. Lange Medical Books/McGrawHill, 2004.
- [5] Shaffer JP. Multiple Hypothesis Testing. *Annu. Rev. Psychol.* 1995. 46:561-84.
- [6] 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.
- [7] Benjamini, Y. & Yekutieli, D. (2001) The control of the false discover rate in multiple testing under dependency. *The Annals of Statistics*. 29(4), 1165-1188.
- [8] Benjamini, Y., & Yekutieli, D. (2005). False discovery rate?adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association*, 100(469), 717-81. doi:10.1198/016214504000001907
- [9] J.D. Storey. "A direct approach to false discovery rates", *Journal of the Royal Statistical Society, B* (2002), 64(3), pp.479-498.

- 240 [10] Polich J. Updating P300: an integrative theory of P3a and P3b. Clin  
241 Neurophysiol. 2007 Oct;118(10):2128-48