NeuroImage 58 (2011) 148–156

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Optimal detection of functional connectivity from high-dimensional EEG synchrony data

Archana K. Singh [a,*], Hideki Asoh [b], Steven Phillips [a]

[a] Human Technology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki, 305-8568 Japan
[b] Intelligent Systems Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki, 305-8568 Japan

ARTICLE INFO

Article history:
Received 19 January 2011
Revised 3 May 2011
Accepted 21 May 2011
Available online 16 June 2011

Keywords:
Optimal discovery procedure
Phase locking values
Synchrony
Electroencephalography
Visual search
Multiple comparison
Positive false discovery rate

ABSTRACT

Computing phase-locking values between EEG signals is a popular method for quantifying functional connectivity. However, this method involves large-scale, high-resolution datasets, which impose a serious multiple testing problem. Standard multiple testing methods fail to exploit the information from the complex dependence structure that varies across hypotheses in spectral, temporal, and spatial dimensions and result in a severe loss of power. They tend to control the false positives at the cost of hiding true positives. We introduce a new approach, called optimal discovery procedure (ODP) for identifying synchrony that is statistically significant. ODP maximizes the number of true positives for a given number of false positives, and thus offers a theoretical optimum for detecting significant synchrony in a multiple testing situation. We demonstrate the utility of this method with PLV data obtained from a visual search study. We also present simulation analysis to confirm the validity and relevance of using ODP in comparison with the standard FDR method for given configurations of true synchrony. We also compare the effectiveness of ODP with our previously published investigation of hierarchical FDR method (Singh and Phillips, 2010).

Introduction

Brain dynamics is increasingly the focus of neuroimaging studies in recent years (Chialvo, 2010; Dauwels et al., 2010; David et al., 2004; Deshpande et al., 2011; Ponten et al., 2010; Schevon et al., 2007). In particular, there is a greater appreciation for the complex role that synchrony plays in cognitive tasks, and how synchrony varies along spatial, temporal and spectral dimensions. For example, significantly greater frontal-parietal synchrony has been reported at a lower gamma band frequency in the visual search task for conditions emphasizing top–down over bottom–up control of attention in monkeys (Buschman and Miller, 2007) and humans (Phillips and Takeda, 2009).

For EEG, computing phase-locking values (PLVs) has become a popular measure for quantifying functional connectivity in terms of synchronization between brain regions (Lachaux et al., 1999). PLVs are computed by wavelet decomposition of EEG signals, providing instantaneous measures of phase differences between two signals at any desired frequency. The advantage of this method over methods based on correlation, covariance, or spectral coherence is that PLV measures are directly applicable to non-stationary signals and treat phase and amplitude independently. Hence, PLV allows one to address more specific questions pertaining to when, where, and at what frequencies do synchronies occur.

Although PLVs permit one to ask more detailed questions pertaining to the nature of brain synchrony, the accompanying increase in data dimensionality raises a serious multiple testing issue. As the distributions of neural events are not uniform and are likely to be dependent on time, frequency and location in the brain, PLV data has a dependence structure, that varies in each dimension. The conventional methods for controlling family-wise error rate (FWER) and false discovery rate (FDR) do not fully exploit the information from this dependence structure, which results in a severe loss of power. They evaluate each hypothesis as a single test ignoring the data and results from across other tests. Therefore, they tend to prevent false positives (Type I error) at the cost of increased false negatives (Type II error). An increase in Type II error leads to missed discoveries, i.e., it fails to detect even the true effects. Although FDR based methods are known to provide a better balance between false positives and false negatives than FWER based methods, dependent and multi-dimensional data attenuate this advantage.

The problems of identifying statistically significant functional connectivity in high-dimensional space–time–frequency synchrony data have been reported by various authors in neuroimaging literature over the last decade (Bhattacharya and Petsche, 2005; Fingelkurts et al., 2003; Kitajo et al., 2010; Krieg et al., 2011; Razoumnikova, 2000; Rodriguez et al., 1999; Singh and Phillips,

* Corresponding author.
E-mail addresses: archana@ni.aist.go.jp, sine.arc@gmail.com (A.K. Singh).
URL: http://staff.aist.go.jp/archana.singh (A.K. Singh).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.05.082

2010; von Stein et al., 1999; Weiss and Rappelsberger, 2000). There is a need to change the current approach for identifying the significant synchrony from PLV data with a complex multi-dimensional dependence structure. Standard approaches to controlling FDR consider all hypotheses as belonging to a single family, and the significance of each hypothesis is evaluated using a t-statistic, or F-statistic that was originally designed for single-hypothesis tests. Most FDR methods extend single-hypothesis testing to multiple-hypotheses testing by combining all hypotheses into one family. An FDR cutoff is then determined based on p-values that are calculated from each point test individually, ignoring information from the other tests. Both steps in this approach fail to exploit information related to dependence structure.

This problem was first identified in context of fMRI imaging of brain activity by Chumbley and Friston(2009) and Chumbley et al. (2010), who advocated an approach based on random field theory. Inference using random field theory is about topological features of the SPM, e.g., number of regions, their spatial extent, or peak height, and thus explicitly accounts for the dependence in neuroimaging data (Worsley et al., 1992). Although the concept of topological FDR has been introduced in context of EEG time–frequency analysis, the method is yet to be applied to an EEG dataset (Kilner and Friston, 2010). For sparsely sampled EEG space–time–frequency data, the topological inference requires a priori knowledge about the region of interest in at least one dimension. This restricts the potential application of topological FDR to either space–time analysis, when the frequency band of interest is known a priori, or time–frequency analysis, when the sensor of interest is known a priori (Kilner et al., 2005; Kilner and Friston, 2010). The computational and conceptual reasons behind this limitation are detailed in Kiebel and Friston (2004).

There are at least two alternatives that avoid the difficulties with topological FDR just mentioned. One way is to organize hypotheses into a hierarchy of families. This approach is employed in the *hierarchical false discovery rate* (hFDR) procedure (Yekutieli, 2008). Another way is to modify the test statistic so that it explicitly shares the dependence information from across all the tests. This approach was suggested by Storey(2007) in his *optimal discovery procedure* (ODP). We outline these two approaches.

The problem of testing multiple hypotheses that are naturally organized into a hierarchical structure is ameliorated by the hFDR procedure (Yekutieli, 2008). A simplified scenario illustrates its advantages over single-family hypothesis testing. If two tests are independent, then the effective significance threshold of either test result is halved, because the chance of exceeding the threshold for declaring a discovery in the absence of a true effect is doubled. If two tests are dependent, in the sense of being identical, then halving the significance threshold is unduly conservative, because the same test is counted twice. Although we do not know in advance precisely which tests are dependent, domain knowledge, or the sorts of hypotheses we wish to entertain help determine a hypothesis-tree hierarchy. Suppose synchrony is known to localize within four frequency bands. Arranging 20 frequency-specific PLV tests into four families (one for each band) and treating each family as a single test explicitly assumes that each test within a family is dependent. Hence, our adjustment for effective significance of each test is reduced from 20 to four, thus raising the sensitivity of our testing procedure. For each family-level test that exceeds threshold, the procedure is recursively applied. For example, we may also wish to test the more detailed hypothesis of synchrony within a particular time band (within a particular frequency band). Since the frequency-band hypotheses that failed to reach threshold are no longer considered, sensitivity is also raised at this second level in the hierarchy. In practice, we found that hFDR was effective in detecting true effects without raising the level of false discoveries beyond the expected level. This method could detect significant instances of synchrony that were consistent with the previously reported findings in the case of the real EEG data and with the true synchrony in the case of simulation (Singh and Phillips, 2010).

A potential difficulty with hFDR occurs when there is no natural, or a priori reason for arranging multiple hypotheses into a particular hierarchy. An alternative method, ODP (Storey, 2007) does not require hierarchically organized families of hypotheses. Instead of employing a single hypothesis test, where information pertaining to the other hypotheses is ignored, ODP uses a statistic designed to simultaneously use information obtained from all other hypotheses for testing the current hypothesis. ODP is defined as the rule that maximizes the number expected true positives (ETP) for each fixed number of expected false positives (EFP), and it also guarantees the optimal FDR (Storey, 2007). The optimizing property of ODP comes from the fact that it is designed to take into consideration the dependence structure among the expected true effects, thus explicitly preventing both false positives and false negatives. An important advantage of using ODP is that it assigns a direct significance measure of FDR to each of the test, eliminating the need to determine a cutoff.

The formulation of the ODP statistic requires estimation of true null and alternate probability densities of the data, which are unknown and vary across application domains. Hence, each application of ODP needs to be developed with careful investigation and consideration to the specific research domain. In this article, we develop an application of ODP for analysis of brain synchrony data obtained from EEG. We show how the ODP statistic can be estimated using the general methodology described in Storey et al.(2007), and evaluate the effectiveness of this procedure in the context of high-dimensional EEG phase locking data. The specificity and sensitivity of ODP is evaluated using simulations, and demonstrated using the experimental data from a visual search study. ODP provides a significant increase in the number of detected true positives (i.e., synchronized electrode pairs for specific times and frequencies) while maintaining FDR compared to standard FDR methods. We also evaluate the effectiveness of ODP in comparison to hFDR and discuss their differences.

Materials and methods

This section includes an overview of the concepts related to ODP and the steps involved in implementing the procedure. A brief overview of hierarchical FDR is also included in the end for the purpose of comparing ODP with hFDR. The details for computing PLV (Lachaux et al., 1999) and with specific application to visual search (Phillips and Takeda, 2009) are not repeated here.

False discovery rate (FDR)

The *false discovery rate* method introduced by Benjamini and Hochberg(1995) (FDRBH) has become a standard method for controlling Type I error in applications involving multiple testing of brain imaging data (Chumbley and Friston, 2009; Genovese et al., 2002; Hemmelmann et al., 2005; Singh and Dan, 2006). FDR is defined as the expected proportion of falsely rejected hypotheses among the rejected ones, which is zero if there are no discoveries: $FDR = E(V/R|R>0)Pr(R>0)$ (see Table 1). FDRBH, like most other

Table 1
Variables associated with the number of true negatives (TN), false negatives (FN), false positives (FP) and true positives (TP), for multiple testing of m null hypotheses.

	Declared Non-significant	Declared Significant	Total
True H_0	U (TN)	V (FP)	m_0
False H_0	T (FN)	S (TP)	$m - m_0$
Total	m-R	R	m

conventional multiple testing methods, follows a *fixed error rate* approach, where the error rate is the desired level of FDR, which determines the rejection region as follows. Let p_i and α denote the ordered p-values and the pre-specified error rate, respectively. Then the rejection region for controlling FDR can be determined as $\gamma = max\{p_i : p_i \leq \alpha \frac{i}{m}\}$, where all null hypotheses corresponding to $p_i \leq \gamma$ are rejected. This method controls FDR at level $\pi_0\alpha$, where $\pi_0 = m_0/m$ is the proportion of true null hypotheses. Therefore, when all the null hypotheses are true ($m_0 = m$), FDR is controlled at level α, and when some of the null hypotheses are rejected ($m_0 < m$), the procedure controls FDR at a level far below α. The power of an FDR controlling procedure can be improved by substituting π_0 with an estimate, $\hat{\pi}_0$ (Benjamini et al., 2006; Storey, 2002).

Positive false discovery rate (pFDR)

Storey(2002) introduced an alternative measure, *positive false discovery rate* (pFDR), conditional on there being positive findings, i.e., at least one discovery: $pFDR = E(V/R|R>0)$ (see Table 1). Instead of the fixed error rate approach, Storey's pFDR method follows a *fixed rejection threshold* approach. The rejection region for this procedure (defined next) is fixed, and then FDR and pFDR are estimated by incorporating $\hat{\pi}_0$.

The procedure for determining the rejection region, which assumes that all null hypotheses are identical with an identical region Γ, is based on observed p-values, $p_1,\ldots, p_m$ for all tests. Let $\Gamma = [0, \gamma]$, where $\gamma \in [0, 1]$, then we reject all the null hypotheses with p-values less than γ. Since p_i are uniformly distributed and $\pi_0 m$ p-values are expected to be null, a conservative estimate of π_0 can be given as follows:

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_i > \lambda\}}{(1-\lambda)m} \tag{1}$$

where, λ is a tuning parameter, $0 \leq \lambda \leq 1$, and # indicates the number of times the condition within parentheses holds true. The optimal choice of λ is determined from a set of possible cutoff values, e.g., $R = \{0, 0.05, \ldots, 0.95\}$, by selecting the value that minimizes the mean squared error over the choice of possible λ, $\widehat{MSE}(\lambda)$, based on an algorithm suggested by Storey(2002) (see steps 2 to 5 in Appendix A).

The total number of rejected hypotheses can be computed as, $R(\gamma) = \#\{p_i \leq \gamma\}$, and the total number of the null hypotheses that are rejected as, $V(\gamma) = m\gamma\pi_0$. Incorporating $\hat{\pi}_0$, FDR and pFDR can be estimated as follows.

$$\widehat{FDR}(\gamma) = \frac{m\gamma\hat{\pi}_0}{max(R(\gamma), 1)} \tag{2}$$

$$\widehat{pFDR}(\gamma) = \frac{\widehat{FDR}(\gamma)}{Pr(R(\gamma) > 0)} = \frac{m\gamma\hat{\pi}_0}{(1-(1-\gamma)^m)max(R(\gamma), 1)} \tag{3}$$

The storey's pFDR method is more powerful than FDRBH, because it directly measures pFDR by incorporating the information of the true null and alternative hypotheses by fixing the rejection region. However, fixing the rejection region may not always be convenient. A more flexible and useful method would be to provide each test with a measure of significance that can be easily interpreted. This is accomplished by q-values, which were introduced as part of pFDR controlling procedure in Storey(2002). A q-value is a pFDR analog of the p-value. It is estimated by calculating the minimum estimated pFDR from among all thresholds, γ, at which the test is called significant from a set of observed p-values (see Eq. (8) in context of ODP). The algorithm for computing q-values is covered in Appendix A. An R package for applying pFDR and computing q-values is also available (http://genomics.princeton.edu/storeylab/q-value/).

Optimal discovery procedure

The core concept of ODP is based on Neymann–Pearson (NP) lemma, which provides a basis for constructing an optimal test statistic for a single testing situation with the observed data (Neyman and Pearson, 1933). Given the observed data, $x = x_1, x_2,\ldots, x_n$ the optimal single test statistic can be defined as a likelihood ratio test, as follows.

$$S_{NP}(x) = \frac{prob(x|H_1)}{prob(x|H_0)}, \tag{4}$$

According to NP lemma, for a given significance level, α, if above likelihood ratio exceeds a given cutoff, say λ, then H_0 can be rejected in favor of H_1. The above NP test is optimal because it is the most powerful among other tests for each fixed Type I error rate. ODP is an extension of NP lemma to facilitate simultaneous testing of multiple hypotheses. ODP reformulates the S_{NP} statistic as the ratio of the sum of the probability of a given test's data under each alternative distribution to that under each null distribution (Eq. (5)). The algorithm for applying ODP involves three steps, which are described below in the context of our EEG application:

1. Computing ODP statistics: Let f_j and g_j represent the respective true null and alternative densities for the jth hypothesis, $j \in \{1,\ldots, m\}$, for the observed PLV data, x_i, from the $i^{th}(i = 1,\ldots, m)$ electrode pair. Let m_0 be the number of true null hypotheses. ODP statistic can be computed as

$$S_{ODP}(x_i) = \frac{\sum_{j=m_{0+1}}^{m} g_j(x_i)}{\sum_{j=1}^{m_0} f_j(x_i)} \tag{5}$$

The data x_i for i^{th} electrode pair is evaluated at the estimated probability density functions for all electrode pairs. As true alternative and null densities are unknown, they can be estimated using the observed data. For example, under the assumption that PLVs follow a normal distribution, the probability densities can be computed using means and variances of the observed data. Substituting the normal density function $\phi(.|\mu, \sigma^2)$ with the parameters μ and σ^2 in Eq. (5),

$$S_{ODP}(x_i) = \frac{\sum_{j=m_{0+1}}^{m} \phi\left(x_i|\mu_{j1}, \sigma_{j1}^2\right)}{\sum_{j=1}^{m_0} \phi\left(x_i|0, \sigma_{j0}^2\right)} \tag{6}$$

where $(\mu_{j1}, \sigma_{j1}^2)$ and $(0, \sigma_{j0}^2)$ are the mean and variance under alternative and null distributions, respectively.

The intuition behind ODP is that the relative significance of each observed x_i increases when multiple true positive results are likely to have similar values. For example, if x_i corresponds to a true alternative hypothesis, then its density, $\phi(x_i|\mu_{i1}, \sigma_{i1}^2)$ will make a substantial contribution to $S_{ODP}(x_i)$. Furthermore, if there are other true alternatives with $\mu_{i1} \approx \mu_{j1}$, then the likelihood of x_i under $\phi(x_i|\mu_{j1}, \sigma_{j1}^2)$ will also make a substantial contribution to $S_{ODP}(x_j)$. It makes sense to increase the significance of a test if there exist other tests with the similar results to maximize ETP for each fixed EFP.

2. Computing bootstrap null distribution: The null distribution of an ODP statistic can be generated by using the following bootstrap procedure. For each i^{th} pair, compute $x_{0i} = x_i - \mu_i$, and construct a bootstrap sample x_i^{0b} by randomly drawing n observations with replacement from x_{0i}. Repeating this B times ($b = 1,\ldots, B$) for each i^{th}

($i=1,\ldots, m$) pair and computing the ODP-statistic from x_i^{0b} generates the bootstrap null distribution of $S_{ODP}(x_i)$. The p-value can be computed from this null distribution as follows:

$$p_i = \frac{\sum_{b=1}^{B}\sum_{j=1}^{m}\left(S_{ODP}\left(x_j^{0b}\right) \geq S_{ODP}(x_i)\right)}{mB} \quad (7)$$

Note in the above equation that the p-value for each i^{th} test takes in to consideration the information from across all $j(=1,\ldots, m)$ tests.

3. Computing q-values for ODP: ODP statistics can be directly thresholded by rejecting all the tests with $S_{ODP}(x_i)\geq\lambda$ for a pre-specified significance cutoff, λ, without computing q-values. However, when fixing a cutoff beforehand is inconvenient, computation of q-values provides a flexible alternative (Storey, 2002, 2003). The q-value of a test measures the minimum false discovery rate that is incurred when calling that test significant. The q-values can be estimated from the p-values and substituting $\widehat{pFDR}(\gamma)$ from Eq. (3) as follows:

$$\hat{q}(p_i) = \min_{\{\gamma \geq p_i\}} \widehat{pFDR}(\gamma) \quad (8)$$

For computing $\widehat{pFDR}(\gamma)$, we need an estimator for π_0, which in turn determines the optimal choice for the cutoff, λ, as described in the detailed algorithm for computing q-values (Appendix A). A q-value threshold is equivalent to directly thresholding the ODP statistic, $S_{ODP}(x_i)\geq\lambda$. Using this direct threshold, $ETP(\lambda)$ and $EFP(\lambda)$ can be estimated from the bootstrap distribution of ODP statistic, and $FDR(\lambda)$ can be determined in terms of these estimates. Theoretical derivations of the proof can be found in Storey et al.(2007).

Note that in the above equation, pFDR estimates are used to compute the q-values. It is possible to compute q-values using FDR estimates in similar terms, i.e., by replacing $\widehat{pFDR}(\gamma)$ with $\widehat{FDR}(\gamma)$ from Eq. (2). However, these FDR analogue of p-values are not robust, as they converge to zero for small p-values, making the inference unreliable (Storey, 2002).

A GUI-based tool for ODP can be downloaded for an implementation in R (http://www.genomine.org/edge/). An R ODP script is also available with the EDGE package to run in it a batch mode. EDGE implementation of ODP generates estimates of q-values and π_0 for a given FDR level. ODP shows the additional information on q-values, such as π_0 over a range of λ and detected discoveries for each q-value cutoff in q-value plots (see Fig. 2 for an example).

Hierarchical FDR

The detailed procedure for hierarchical FDR (hFDR) is described in Yekutieli(2008), and in the context of PLV analysis in Singh and Phillips(2010). The hFDR procedure is implemented by organizing the hypotheses into a family-subfamily tree hierarchy, where each (sub) family is associated with a single hypothesis. For instance, the tests for PLV data are grouped into M frequency and N time families based on the frequency and time band associated with the test. The M frequency families constitute the first level of the hierarchy. In this case, there are N time subfamilies at the second level for each frequency family, and within each time subfamily are the test statistics, one for each electrode pair, at the third (lowest) level. Alternatively, the hierarchy may include the time families at the first level, and the frequency families at the second level, and the hypotheses for the electrode pairs at the third level. The data associated with each test within a family is summarized (i.e., averaged) to become the data for that family and its associated single hypothesis. The testing begins at the first level by applying a single-sample t-test and FDRBH control to test the M hypotheses in frequency family. If any hypothesis is rejected, then testing continues by testing the corresponding time subfamily at the second level. This process continues by recursively checking each child hypothesis of a parent that was rejected, and terminating upon not rejecting any children. The FDR bound on a hypothesis tree is defined recursively as the sum of the expected proportion of the number of false discoveries to total discoveries for each family. Yekutieli (2008) derived an approximate bound

$$bound = q\ \delta\ \frac{N_d + N_f}{N_d + 1}, \quad (9)$$

where N_d is the number of observed discoveries, N_f is the number of families tested, and δ, a multiplicative constant, is set to 1. This bound varies in an interval $[q, 2q]$, where q is the expected FDR level. When the number of discoveries far exceeds the number of family tested, the FDR bound in Eq. (9) converges to q. Note that q as defined here should not be confused with q-value, which is defined as a significance measure of FDR in Optimal Discovery procedure section.

Simulation

Simulations were performed to access the specificity and sensitivity of ODP for multidimensional EEG data. As in our previous study, true effects were assigned to particular frequency–time bands. Hence, we associated PLV differences (between conditions) with 2 frequency bands, 12 time bands, 10 participants, and 25 electrode pairs (i.e., all pairwise combinations of 5 frontal and 5 parietal electrodes) constituting a $2\times12\times10\times25$ array. For frequency–time windows containing significant effects, 10 out of 25 pairs were defined as truly significant. The proportion of electrode pairs with true synchronies, π_1, varied from 8% to 48% over a range of time–frequency windows. To induce dependence across the time windows for specific frequency bands, the PLVs for their corresponding electrode pairs were generated from a multivariate normal distribution with the parameters μ and $\sum=\sigma^2R$. The mean vector, μ was assigned zero PLV effects for true null hypotheses, and positive PLV effects to represent the true effects pertaining to alternative hypotheses. The covariance matrix, $\sum=\sigma^2R$ was constructed by assigning the variance, σ^2 and correlation matrix, R from a real PLV dataset. The tuning parameter λ was automatically chosen using a bootstrap distribution with 100 resamples following Storey(2002) (refer to Appendix A for the algorithm).

ODP explicitly maximizes the number of expected true positives for a given level of FDR. Hence, we were also interested in comparing ODP and hFDR methods on grounds of the false positives and false negatives that they incur. A direct comparison between ODP and hFDR is difficult as each works on a distinct operating principle, using a different measure of false positives. Nevertheless, a comparative evaluation of their detection power and actual FDR incurred with synthetic data, where configurations of true discoveries are known, would be useful for exploring any specific conditions that may warrant the use of one method over the other. For hFDR analysis, we constructed a 3-level hierarchical FDR tree so as to cast the hypotheses belonging to the frequency dimension at the first level, those belonging to the time dimension at the second level, and to each electrode pair at the third level. A 5% FDR threshold was used in ODP, hFDR, and FDRBH.

The simulations were performed in R, and ODP was applied using the R script. Reported numbers of detected discoveries, false positives, false negatives are averages over 100 runs, for both ODP and hFDR applications. In the case of ODP, we also reported the estimated π_0 for each case of assigned true positives.

Experimental data analysis

To confirm the effectiveness of ODP on real data, we reanalyzed EEG data acquired from a visual search experiment (Phillips and

Takeda, 2009). The purpose of the experiment was to test the hypothesis that top–down driven control of visual attention in humans is accompanied by frontal-parietal synchrony in the lower gamma-band. Top–down signals were induced using distractors that share a feature (e.g., color, or orientation) with the target, yielding a steep search slope (search time increasing with display set size)—inefficient search. Bottom–up signals were induced using distractors with no feature in common with the target, yielding a flat slope (search time independent of set size)—efficient search. Participants showed significantly greater synchrony between frontal and parietal electrodes in the lower gamma-band during inefficient than efficient search, which replicated the same effect observed in monkeys (Buschman and Miller, 2007). Our region of interest was confined to 25 electrode pairs, i.e., five frontal electrodes (F7, F3, Fz, F4, F8) by five posterior electrodes (T8, P3, Pz, P4, T6), located according to the International 10–20 system. Frequency was partitioned into lower (22–34 Hz) and upper (36–48 Hz) gamma bands, corresponding to the original studies (Buschman and Miller, 2007; Phillips and Takeda, 2009; Singh and Phillips, 2010). Time was partitioned into twelve 50 ms windows for the first 600 ms after stimulus (search display) onset. ODP was performed using R EDGE software. The estimation of π_0 was done with bootstrap option. The FDR level was specified as 10% and $\lambda = [0, 0.05,\ldots, 0.95]$.

For FDRBH, all $600 (= 2 \times 12 \times 25)$ hypotheses were regarded as a single family, the resulting p-values were thresholded by FDRBH at a level of 10%.

For hierarchical FDR, we used a 3-level tree, with 2 hypothesis for each frequency band at the first level, 12 hypotheses for the time bands at the second level, and 25 electrode pair hypotheses corresponding to each time band at the third level. See Singh and Phillips(2010) for detailed application of hFDR. The results from ODP and hFDR were overlaid on synchrony maps using matlab. The level of FDR was set at 10% at each hierarchical level.

Results

Simulation

The main purpose of performing the simulations was to evaluate how well ODP extracts the true effects while maintaining FDR, within a pre-defined FDR level, at given proportions of true synchronies under typical dependence structure of PLV data from EEG. In addition, we also wanted to compare the performance of ODP against hFDR and FDRBH when the ground truth was known. The simulation results are summarized in Fig. 1. ODP exhibited a linear increase in the number of false positives, a linear decrease in the number of false negatives, and a linear increase in FDR as the proportion of true discoveries, $(\pi_1 = 1 - \pi_0)$, increased. FDR was maintained within the specified level of 5%, except when proportion of true discovery was 48%. For hFDR, there was a linear increase in both the number of false positives, and FDR, which was below 5% in all the cases. hFDR detected more discoveries than ODP when the proportion of true discoveries was less than 20%. In the case of FDRBH, fewer discoveries were detected than ODP. In comparison to ODP, FDRBH registered higher levels of false positives and FDR for lower proportions of true discoveries, i.e. when $\pi_1 \leq .24$ and higher levels of false negatives for all proportions. The level of FDR in FDRBH remained more or less constant within the specified cutoff. The estimates of the proportion of null hypotheses, $\hat{\pi}_0$ were lower than the proportion of assigned null hypotheses $(\pi_0 = 1 - \pi_1)$ for all simulations (see Table 2).

Experiment

ODP failed to detect any significant discoveries at 5% FDR level, but showed 89 discoveries at 10% FDR level (Fig. 3). The figure shows significantly greater synchrony for the inefficient than efficient condition (red lines) in the lower gamma band (22–34 Hz), predominantly over 250–550 ms post-stimulus time interval. There was not much evidence for significantly greater synchrony for the inefficient than efficient condition for the high gamma band (36–48 Hz), and for significantly greater synchrony in efficient than inefficient condition (blue lines) in either of the bands. In ODP analysis, we also estimated q-values and $\hat{\pi}_0(\lambda)$ at $\lambda = (0, 0.05,\ldots, 0.95)$ as shown in q-value plots in Fig. 2.

For the purpose of comparison, hFDR detected 26 significant pairs at 5% FDR level as reported in Singh and Phillips(2010), and 45 significant pairs at the 10% FDR level, where the detections were predominantly over 300–500 ms post-stimulus interval (Fig. 4). There was no detection of significantly greater synchrony for the inefficient than efficient condition neither at the high gamma band (36–48 Hz), nor for significantly greater synchrony in efficient than inefficient condition in either bands. FDRBH showed only 10 pairs with significantly greater synchrony in the lower gamma band for

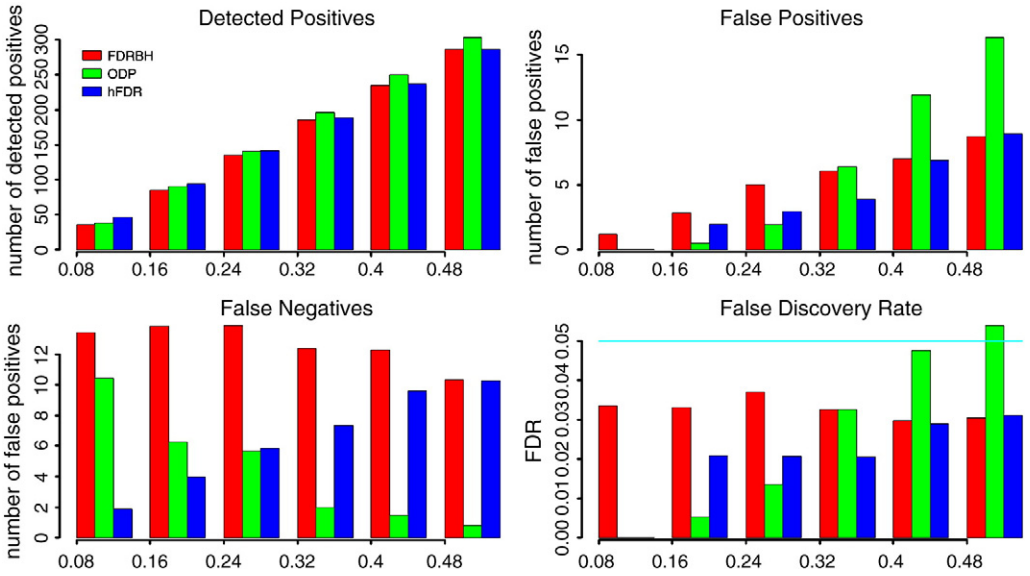

Fig. 1. Comparison between BHFDR, hFDR, and ODP. Each T-F window with assumed true synchrony effects has 10 pairs with true synchrony pairs; the x-axis represents the proportion of true positives; the graph panels represent: the total number of detected positives (top-left), the total number of false positives (top-right), the total number of false negatives (bottom-left), and false discovery rate (bottom-right).

Table 2
True vs. estimated π_0 in simulation.

Assigned π_1	Assigned π_0	Estimated π_0 (SD)
0.08	0.92	0.70 (0.016)
0.16	0.84	0.68 (0.050)
0.24	0.76	0.72 (0.036)
0.32	0.68	0.50 (0.040)
0.40	0.60	0.45 (0.030)
0.48	0.52	0.39 (0.038)

inefficient than efficient condition at 10% cutoff (results for FDRBH are not shown).

Discussion

This article describes a novel application of optimal false discovery rate procedure as an answer to the problem of detecting statistically significant synchrony in multiple testing of EEG phase locking analysis. EEG phase locking analysis involves a complex dataset, with a small proportion of significant synchrony effects in multiple dimensions of time, frequency, and electrode space. The dependence structure of phase locking values may vary within each dimension. Most existing FDR procedures, including FDRBH, are implemented by thresholding p-values that are obtained individually from single hypothesis tests, e.g., a *t*-test, in such a way that it prevents the dependence structure among the expected true discoveries to be incorporated in the simultaneous evaluation of significance from multiple hypotheses testing. Therefore, they tend to prove too conservative to reveal even true significant discoveries. This problem has been discussed by several authors over the past decade, who either resorted to a non-confirmatory functional connectivity analysis or opted for a smaller region of interested based on previous research. ODP overcomes this problem by minimizing missed discoveries (false negatives) for each fixed FDR level. There are two key aspects behind this optimization of power.

First, the ODP statistic is defined as a function of both expected numbers of false positives and false negatives, which are estimated from the data itself. Second, ODP applies a simultaneous thresholding procedure to all the tests, where each test is evaluated by including the relevant information from all the tests. This is evident from Eqs. (5) and (6) for computing the ODP statistic and in Eq. (7) for determining its significance. Therefore, if data x_i shares any similarity in its structure with data x_j, then its density function under the alternative null hypothesis will contribute substantially to $\hat{S}_{ODP}(x_i)$ (Eq. (6)). Thus, the estimation of ODP for each hypothesis includes the dependence information from the entire set of hypotheses. In addition to sharing commonality, each ODP statistic also takes into account the variance structure within each electrode pair. The electrode pairs with smaller variance in PLV have greater chance to contribute to $\hat{S}_{ODP}(x_i)$.

Our simulations showed that ODP optimized the FDR as a linear function of the proportion of true discoveries. When the proportion of true discoveries was less than 48%, ODP controlled FDR at the desired 5% level (Fig. 1). We found that when the proportion of true discoveries was 80%, FDR increased to 8%. Although FDR exceeded the desired 5% level for high proportion true effects (i.e., greater than 48%), such situations are unlikely to occur in real neuroimaging data, where the true effects are more likely to be sparsely distributed across the regions tested. This questions the practice of fixing the FDR level in advance without taking in to consideration the expected proportion of true effects in the data, which differs in each application domain. The conventional threshold for FWER controlling procedures is set at 5%. For FDR, there are no such standards, though most existing FDR controlling methods impose the need to fix an acceptable FDR level before any data are seen. ODP eliminates this requirement and offers a more flexible interpretation of FDR by incorporating a q-value based measure of significance for FDR. We may report the q-value for every electrode pair and let researchers choose a level of desirable significance. A q-value plot, that is available in ODP's EDGE implementation in R, serves as a reference guide for selecting an appropriate FDR cutoff by showing the number of significant discoveries for the selected range of FDR cutoffs, e.g., see Fig. 2 for experiment data analysis. The minimum q-value is 0.075, which revealed significant synchrony for 46 electrode pairs. Choosing a FDR cutoff of 10% reveals significant synchrony in 89 electrode pairs. In the case of FDRBH, it fails to produce significant evidences of synchrony at

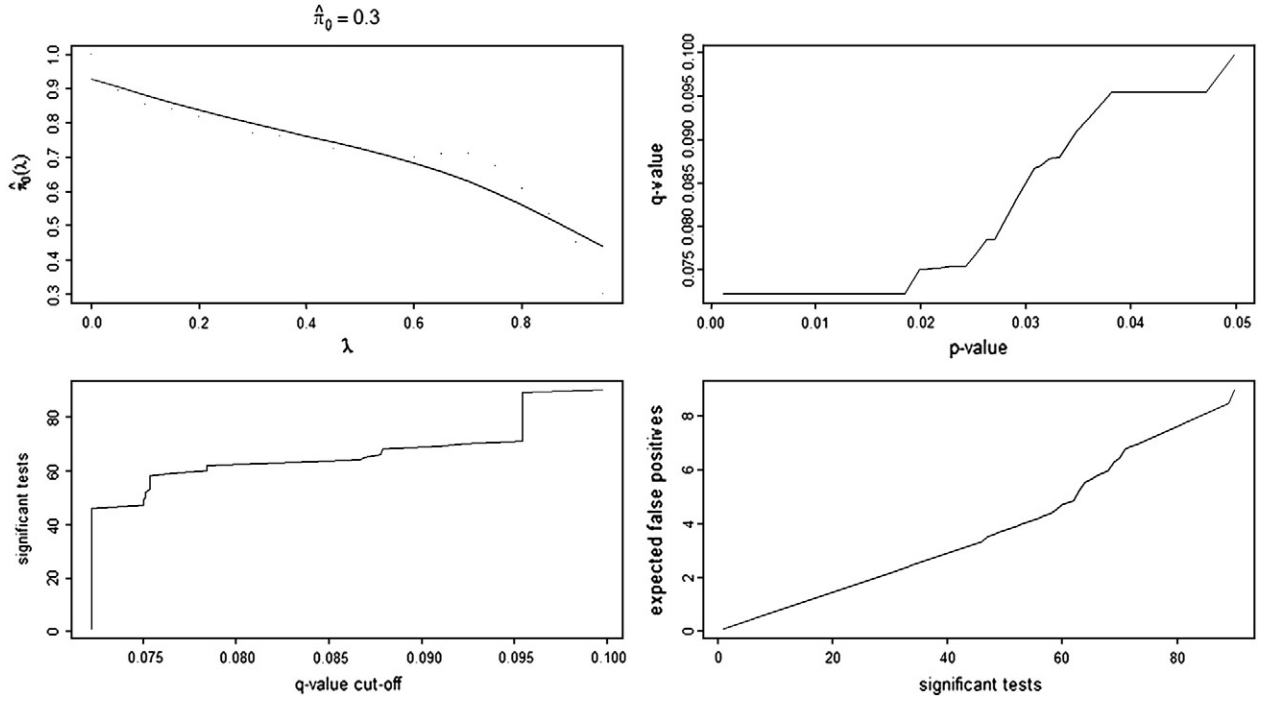

Fig. 2. Q-value plots generated by ODP for experimental data.

this level. In comparison with ODP, FDRBH showed much higher levels of false negatives, and tends to be more conservative as the proportion of true discoveries increases.

Our simulations and experimental data pertained to just 19 electrodes, and yet the problem of multiple comparisons was severe. The growing popularity of high-density (e.g., 64- to 256-channels) EEG clearly exacerbates the multiple comparisons problem. We first advocated the need to change the existing approach to controlling false positives with complex and high-dimensional PLV data in our previous study and proposed using a hierarchical approach for controlling FDR (Singh and Phillips, 2010). hFDR method follows a hierarchical testing scheme where PLV effects in given instances of frequency, time, and electrode pairs are tested at different levels in the hierarchical hypothesis tree. The test data is averaged within each dimension to split the hypotheses into hierarchically organized multiple families. This implicitly incorporates the dependence structure within each dimension. In this way, multiple smaller families are more likely than a single large family to survive FDR correction. An overall FDR can be computed by the summing over the FDRs of all the families that are tested along the hierarchical tree.

hFDR could detect 26 pairs with significant synchrony in our experimental analysis at 5% overall FDR level, whereas ODP failed to detect any synchrony at this FDR level. This result with hFDR was published in Singh and Phillips(2010). On the other hand, when we raised the expected FDR level to 10%, ODP could detect approximately twice as many synchronous pairs as did hFDR. In the simulation analysis, we found that the detection power of each method varied depending on the configurations of true synchrony. Our simulations have shown that ODP outperformed hFDR when the proportion of true effects was large. However, when the proportion of the true effect was small, hFDR outperformed ODP by detecting more pairs. This result is consistent with the simulation results of a technical report on ODP, where the performance of ODP has been reported in comparison with other multiple comparison methods (hFDR was not included). The authors observed a tendency of ODP to underperform the other methods when the percentage of differentially expressed genes was small (Storey, 2005). The performance of ODP depends on the homogeneity of the dependence structure among hypotheses. It is expected to perform better with asymmetric data than symmetric data. Asymmetry is implied when most PLV effects are in a particular direction, i.e. they are either mostly positive or mostly negative, which indeed seems to be the case with our experimental data, that showed more significant PLV in inefficient–efficient contrast than in efficient–inefficient contrast (Figs. 3 and 4). The simulated data was also asymetric (with greater proportion of positive PLV effects than negative PLV effects) as it was generated using the parameters obtained from a real PLV dataset. For a comparison, we ran the simulation with an induced symmetry by assuming both positive and negative PLV effects in similar proportions. We observed that symmetry reduced the number of detections by ODP considerably when the proportion of true discoveries was smaller than 25%. Symmetry did not affect the performance of hFDR (results not shown here).

In the experiment analysis, hFDR and ODP offered almost similar qualitative results, showing the most prominent synchrony in the similar post-stimulus time intervals. While the inference obtained from hFDR and ODP are comparable, each method has its own advantages. hFDR approach is more amenable for datasets that are naturally hierarchical and is particularly powerful when the synchrony effects are concentrated in a few families of hypotheses. However, the PLV effects may not be bound to a given dimension (or a few families). In such cases, ODP may be more powerful.

Although the result of ODP has a straightforward theoretical interpretation, each practical application requires some knowledge about the data, specifically for estimating ODP and setting the tuning parameter, λ. The computation of the ODP statistic in Eq. (6) is based on the assumed normal probability density function for the data, which holds reasonably well in the case of PLVs (Doesburg et al., 2008; Lachaux et al., 1999). Note that this assumption will only influence the estimated ODP statistic (as compared to the true ODP statistic). The significance of ODP is computed using non-parametric bootstrap procedure, and therefore the correct form of the distribution need not be known in advance. The selection of λ is important for the accuracy in estimating π_0, and it depends on the distribution of p-values that varies in each application. Fortunately, options are available for automating the optimal value of λ from a range of possible values, namely, the bootstrap option and smoothing spline option as described in Storey(2002, 2003), respectively. In the bootstrap option, the optimal choice of λ is determined from the bootstrap distribution of $\pi_0(\lambda)$ as the value with the least mean square error (MSE). As this

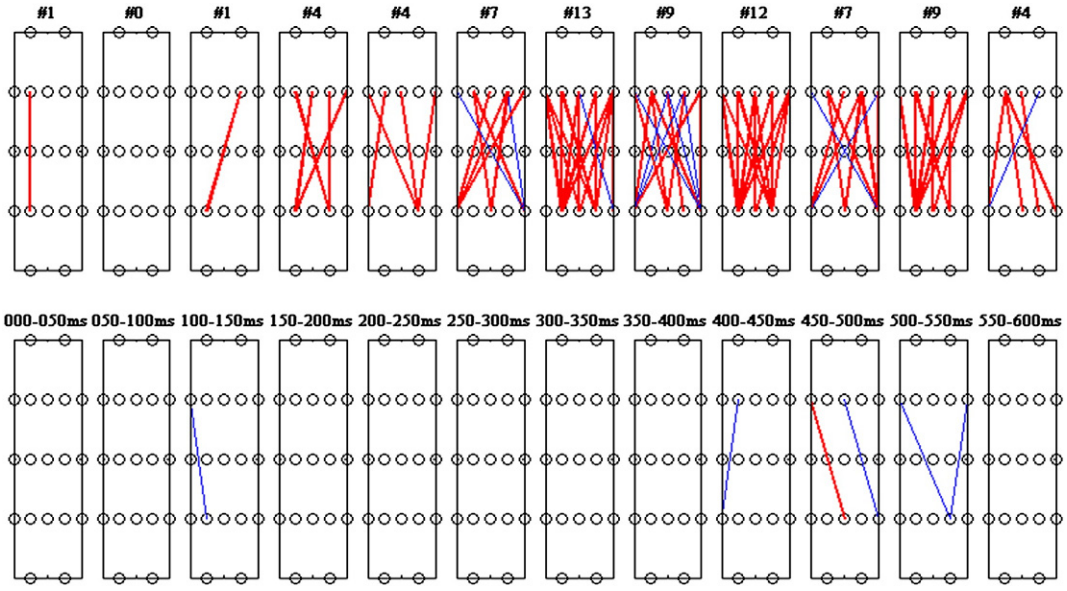

Fig. 3. The synchrony map for experimental data from ODP application indicating (number of) electrode pairs showing significantly greater phase-locking for the inefficient than efficient search conditions (red lines) for 10% FDR level, i.e. $q-value<0.1$. The top and bottom rows correspond to lower (22–34 Hz) and higher (36–48 Hz) gamma bands.

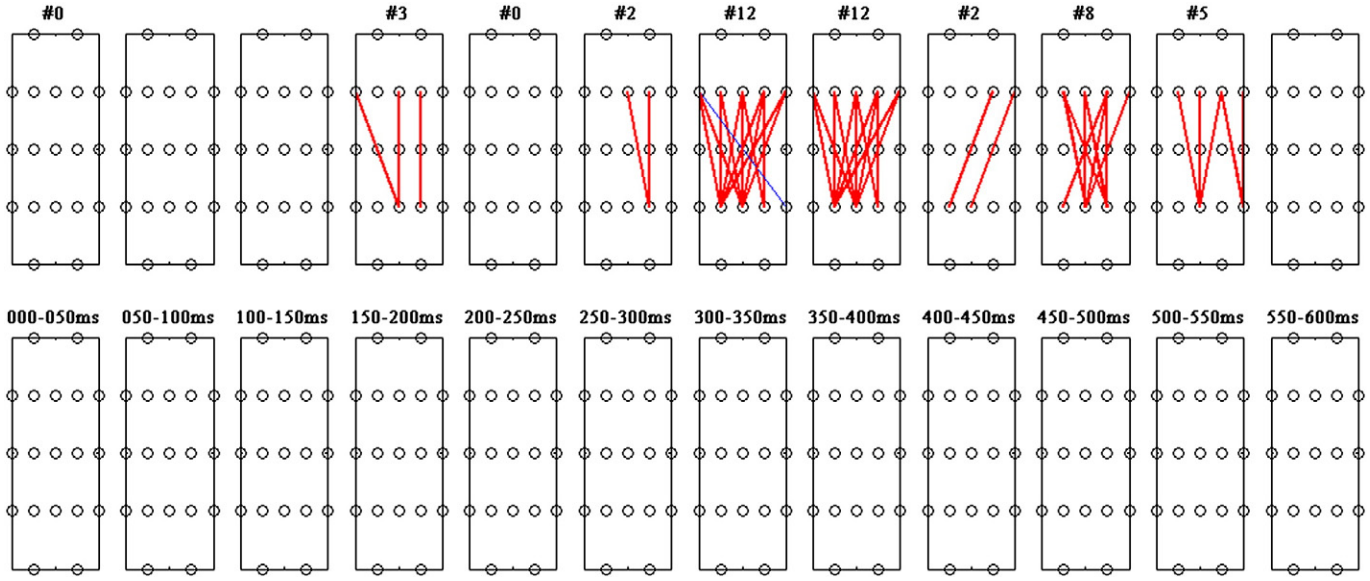

Fig. 4. Synchrony map from hFDR application indicating (number of) electrode pairs showing significantly greater phase-locking for the inefficient than efficient search conditions. The top and bottom rows correspond to lower (22–34 Hz) and higher (36–48 Hz) gamma bands.

option is considered as safe and conservative for any distributional form of p-values, we chose this option for presenting our analysis. In the case of smoothing spline, π_0 is estimated by fitting $\hat{\pi}_0(\lambda)$ for all possible values of λ. The accuracy of this fitting depends on how well the smoothing order is specified. Over fitting may result in inflated results. For a small number of p-values, it may be more conservative than the bootstrap option. For the purpose of comparison, we analyzed our experiment data using both these options. The range of λ was set as specified in the algorithm, and for the smoothing spline option we used 3 degree of freedom following Storey(2003). The q-values from this option were more conservative, and failed to detect any synchrony at 10% cutoff (results are not shown).

The idea of incorporating dependence structure to improve the significance from point test in neuroimaging studies is not new. Worsley et al.(1992) introduced random field theory that controlled FWER to infer on the topological features of SPMs for fMRI images. This concept was later extended to include a topological FDR method (Chumbley and Friston, 2009; Chumbley et al., 2010; Kilner and Friston, 2010). Though these articles discuss the possible implementation of topological FDR for EEG time–frequency data, there is a lack of illustrative applications. A topological inference may offer a limited scope for analyzing PLVs in space–time–frequency domain, especially with sparsely sampled EEG data due to several reasons. The current implementations of topological inference methods, e.g. topological FDR are practically limited to either space–time (when frequency band of interest is known a priori) or time–frequency (when sensor of interest is known a priori) search space but not both. The topological inference for a space–time–frequency PLV analysis needs to be extended to a 4D (space× space×time×frequency) search space, and this has not yet been implemented due to computational and conceptual reasons (Kiebel and Friston, 2004; Kilner et al., 2005).

Furthermore, the sensitivity of a topological inference depends on how well the assumptions required by RFT hold, e.g., sufficiently smoothed data and sufficiently high threshold. There is a lack of clear guidelines on how smooth images should be (Hayasaka and Nichols, 2003; Kilner and Friston, 2010; Kilner et al., 2005). (Kilner et al., 2005) recommend that the size of smoothing kernel should be selected based on the reported variability in the latency and frequency peak in time–frequency effects over subjects, and such information may not always be available. Similarly, there is no consensus on how high the threshold should be for random field theory to work. Holmes(1994) compared different thresholds in a simulation study and found the RF test to be conservative for low thresholds. Poline et al.(1997) found that the RF test is anticonservative for low thresholds and becomes conservative for high thresholds. Kilner et al.(2005) could not detect any significant time-frequency effect in an EEG study, based on the corrected height threshold estimated by the random field theory.

ODP offers a more flexible approach, which does not require any assumptions regarding the smoothness and dependence in the data. A topological inference may be an interesting perspective of the current study. However, it requires a separate investigation of the practical aspects of the existing topological FDR method for analyzing EEG data in space–time–frequency domain.

Another approach that advocated incorporating dependence by sharing common information across the test for reducing Type I and Type II errors was proposed in a method called *local FDR* by Efron and Tibshirani(2002). This approach assumed a hierarchical Bayes model, where the priors of the parameters are common among all tests, and the priors are hierarchically estimated by using the data from all tests. In local FDR, the distribution of the test statistics follows a mixture density model, and the inference is obtained by comparing the density of the null to the mixture distribution for a given statistic. See Schwartzman et al.(2009) for the application of local FDR in context of thresholding statistical parametric maps of fMRI data. ODP and local FDR are equivalent if local FDR is applied using a statistic compatible to ODP statistic, as shown in a microarray study that compared the two procedures (Oba and Ishii, 2009). A future investigation on local FDR and its comparison with ODP would be worth examining in context of multiple testing of PLVs. If local FDR can be adopted to multiple testing of PLVs such that it offers the same optimal level of thresholding as ODP, it will serve as a useful Bayesian alternative to frequentist optimal discovery procedure.

Recent research in brain dynamics has established a number of methods for quantifying functional connectivity in EEG studies. See Dauwels et al.(2010) and David et al.(2004) for a review of some of these methods. Considering that no standard solutions exist to address the issue of multiple comparison in high-dimensional EEG synchrony data that result from these methods, this article can serve as a useful example for other applications too. ODP is readily applicable to the methods, e.g., mutual information, generalized synchronization (Quian Quiroga et al., 2002), single-trial phase locking (Lachaux et al., 2000), structural synchrony (Fingelkurts et al., 2003), empirical mode detection PLV (Sweeney-Reed and Nasuto, 2007) phase resetting (Makinen et al., 2005; Thatcher et al., 2008) and a recent method based on Cohen's class of time–frequency distributions

(Aviyente et al., 2011) specifically in electrode space when MR images are not available.

Acknowledgment

This work was supported by a JSPS Postdoctoral Fellowship to the first author and JSPS grant-in-aid (22300092). We thank Yuji Takeda and Shigeyuki Oba for their advice. We also thank Daniel Yekutieli for his advice on hFDR.

Appendix A. Computing q-values

A q-value is computed as the minimum FDR over which a test statistic can be rejected for the specified range of λ. The following algorithm is summarized from Storey(2002) (for bootstrap option) and Storey(2003) (for smoothing spline option).

1. Let p_i be the ordered p-values.
2. For a range of λ, say $R=\{0, 0.05,\ldots, 0.95\}$, compute $\hat{\pi}_0$ according to Eq. (1). For smoothing spline option, $\hat{\pi}_0 = \hat{f}(1)$, where $\hat{f}$ is a natural cubic spline with 3 df of $\hat{\pi}_0(\lambda)$ on λ, and skip steps 3–5.
3. For each $\lambda \in R$, form bootstrap null distributions $\hat{\pi}_0^{*b}(\lambda)$, for $b=1,\ldots, B$, by taking bootstrap samples of the p-values.
4. For each $\lambda \in R$, estimate its mean square error (MSE) as

$$\widehat{MSE}(\lambda) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\pi}_0^{*b}(\lambda) - min_{\lambda' \in R}\left\{\hat{\pi}_0(\lambda')\right\}\right]^2. \quad (10)$$

5. Set $\hat{\lambda} = arg\ min_{\lambda \in R}\{\widehat{MSE}(\lambda)\}$. Then, the overall estimate of π_0 is $\hat{\pi}_0 = \hat{\pi}_0\left(\hat{\lambda}\right)$.
6. Set $q\left(\hat{p}_m\right) = \widehat{pFDR}(p_m)$, where $\widehat{pFDR}(p_m)$ is obtained from Eq. (3) after incorporating $\hat{\pi}_0$.
7. Set $q\left(\hat{p}_i\right) = min\ \{\widehat{pFDR}(p_i),\ \widehat{pFDR}(p_{i+1})$ for $i=m-1, m-2,\ldots, 1$.

References

Aviyente, S., Bernat, E.M., Evans, W.S., Sponheim, S.R., 2011. A phase synchrony measure for quantifying dynamic functional integration in the brain. Human Brain Mapping 32 (1), 80–93.

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 57 (1), 289–300.

Benjamini, Y., Krieger, A.M., Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93 (3), 491–507 (September).

Bhattacharya, J., Petsche, H., 2005. Phase synchrony analysis of EEG during music perception reveals changes in functional connectivity due to musical expertise. Signal Processing 85 (11), 2161–2177.

Buschman, T.J., Miller, E.K., 2007. Top–down versus bottom–up control of attention in the pr tal and posterior parietal cortices. Science 315 (5820), 1860–1862.

Chialvo, D.R., 2010. Emergent complex neural dynamics. Nature Physics 6, 744–750.

Chumbley, J.R., Friston, K.J., 2009. False discovery rate revisited: FDR and topological inference using gaussian random fields. NeuroImage 44 (1), 62–70.

Chumbley, J., Worsley, K., Flandin, G., Friston, K., 2010. Topological FDR for neuroimaging. NeuroImage 49 (4), 3057–3064.

Dauwels, J., Vialatte, F., Musha, T., Cichocki, A., 2010. A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. NeuroImage 49 (1), 668–693.

David, O., Cosmelli, D., Friston, K.J., 2004. Evaluation of different measures of functional connectivity using a neural mass model. NeuroImage 21 (2), 659–673.

Deshpande, G., Santhanam, P., Hu, X., 2011. Instantaneous and causal connectivity in resting state brain networks derived from functional MRI data. NeuroImage 54 (2), 1043–1052.

Doesburg, S.M., Roggeveen, A.B., Kitajo, K., Ward, L.M., 2008. Large-scale gamma-band phase synchronization and selective attention. Cerebral Cortex 18 (2), 386–396.

Efron, B., Tibshirani, R., 2002. Empirical bayes methods and false discovery rates for microarrays. Genetic Epidemiology 23 (1), 70–76.

Fingelkurts, A., Fingelkurts, A., Krause, C., Kaplan, A., Borisov, S., Sams, M., 2003. Structural (operational) synchrony of EEG alpha activity during an auditory memory task. NeuroImage 20 (1), 529–542.

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15 (4), 870–878.

Hayasaka, S., Nichols, T., 2003. Validating cluster size inference: random field and permutation methods. NeuroImage 20, 2343–2356.

Hemmelmann, C., Horn, M., Ssse, T., Vollandt, R., Weiss, S., 2005. New concepts of multiple tests and their use for evaluating high-dimensional EEG data. Journal of Neuroscience Methods 142 (2), 209–217.

Holmes, A., 1994. Statistical issues in functional brain mapping. Ph.D. thesis, University of Glasgow.

Kiebel, S., Friston, K., 2004. Statistical parametric mapping for event-related potentials i: generic considerations. NeuroImage 22 (2), 492–502.

Kilner, J., Friston, K., 2010. Topological inference for EEG and MEG. The Annals of Applied Statistics 4 (3), 1272–1290.

Kilner, J., Kiebel, S., Friston, K., 2005. Applications of random field theory to electrophysiology. Neuroscience Letters 374, 174–178.

Kitajo, K., Doesburg, M., Yamanaka, K., Nozaki, Ward, L.M., Yamamoto, Y., 2010. Effects of Parkinson's disease on brain-wave phase synchronisation and cross-modulation. EPL (Europhysics Letters) 89 (4), 48001.

Krieg, J., Fonseca, A.T.-D., Martínez-Montes, E., Marquis, P., Liégeois-Chauvel, C., Bénar, C.-G., 2011. A comparison of methods for assessing alpha phase resetting in electrophysiology, with application to intracerebral EEG in visual areas. NeuroImage 55 (1), 67–86.

Lachaux, J., Rodrigue, E., Martnerie, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. Human Brain Mapping 8 (4), 194–208.

Lachaux, J.-P., Rodriguez, E., Quyen, M.L.V., Lutz, A., Martinerie, J., Varela, F.J., 2000. Studying single-trials of phase synchronous activity in the brain. International Journal of Bifurcation and Chaos in Applied Sciences and Engineering 10 (10), 2429–2439.

Makinen, V., Tiitinen, H., May, P., 2005. Auditory event-related responses are generated independently of ongoing brain activity. NeuroImage 24 (4), 961–968.

Neyman, J., Pearson, E.S., 1933. On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London. Series A 231 (10), 289–337.

Oba, S., Ishii, S., 2009. Optimal sufficient statistics for parametric and non-parametric multiple simultaneous hypothesis testing. International Journal of Biostatistics 5 (1).

Phillips, S., Takeda, Y., 2009. Greater frontal-parietal synchrony at low gamma-band frequencies for inefficient than efficient visual search in human EEG. International Journal of Psychophysiology 73 (3), 350–354.

Poline, J., Worsley, K., Evans, A., Friston, K., 1997. Combining spatial extent and peak intensity to test for activations in functional imaging. NeuroImage 5 (2), 83–96.

Ponten, S., Daffertshofer, A., Hillebrand, A., Stam, C., 2010. The relationship between structural and functional connectivity: graph theoretical analysis of an EEG neural mass model. NeuroImage 52 (3), 985–994.

Quian Quiroga, R., Kraskov, A., Kreuz, T., Grassberger, P., 2002. Performance of different synchronization measures in real data: a case study on electroencephalographic signals. Physical Review E 65 (4), 041903.

Razoumnikova, O.M., 2000. Functional organization of different brain areas during convergent and divergent thinking: an EEG investigation. Cognitive Brain Research 10 (1–2), 11–18.

Rodriguez, E., George, N., Lachaux, J.P., Martinerie, J., Renault, B., Varela, F., 1999. Perception's shadow: long-distance synchronization of human brain activity. Nature 397, 430–433.

Schevon, C.A., Cappell, J., Emerson, R., Isler, J., Grieve, P., Goodman, R., McKhann Jr., G., Weiner, H., Doyle, W., Kuzniecky, R., Devinsky, O., Gilliam, F., 2007. Cortical abnormalities in epilepsy revealed by local EEG synchrony. NeuroImage 35 (1), 140–148.

Schwartzman, A., Dougherty, R.F., Lee, J., Ghahremani, D., Taylor, J.E., 2009. Empirical null and false discovery rate analysis in neuroimaging. NeuroImage 44 (1), 71–82.

Singh, A.K., Dan, I., 2006. Exploring the false discovery rate in multichannel NIRS. NeuroImage 33 (2), 542–549.

Singh, A.K., Phillips, S., 2010. Hierarchical control of false discovery rate for phase locking measures of EEG synchrony. NeuroImage 50 (1), 40–47.

Storey, J.D., 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (3), 479–498.

Storey, J.D., 2003. The positive false discovery rate: a bayesian interpretation and the q-value. The Annals of Statistics 31 (6), 2013–2035.

Storey, J.D., 2005. The optimal discovery procedure: a new approach to simultaneous significance testing. UW Biostatistics Working Paper Series, Working Paper 259 http://www.bepress.com/uwbiostat/paper259/ (Available at) http://www.bepress.com/uwbiostat/paper259/.

Storey, J.D., 2007. The optimal discovery procedure: a new approach to simultaneous significance testing. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69 (3), 347–368.

Storey, J.D., Dai, J.Y., Leek, J.T., 2007. The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. Biostatistics 8 (2), 414–432.

Sweeney-Reed, C., Nasuto, S., 2007. A novel approach to the detection of synchronisation in EEG based on empirical mode decomposition. Journal of Computational Neuroscience 23, 79–111.

Thatcher, R., North, D., Biver, C., 2008. Intelligence and EEG phase reset: a two compartmental model of phase shift and lock. NeuroImage 42 (4), 1639–1653.

von Stein, A., Rappelsberger, P., Sarnthein, J., Petsche, H., 1999. Synchronization between temporal and parietal cortex during multimodal object processing in man. Cerebral Cortex 9 (2), 137–150.

Weiss, S., Rappelsberger, P., 2000. Long-range EEG synchronization during word encoding correlates with successful memory performance. Cognitive Brain Research 9 (3), 299–312.

Worsley, K.J., Evans, A.C., Marrett, S., Neelin, P., 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. Journal of Cerebral Blood Flow and Metabolism 12 (6), 900–918.

Yekutieli, D., 2008. Hierarchical false discovery rate controlling methodology. Journal of the American Statistical Association 103 (481), 309–316.