Journal of Neuroscience Methods 268 (2016) 14–22

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Robust power spectral estimation for EEG data

Tamar Melman[a,*], Jonathan D. Victor[b]

[a] Tri-Institutional Program in Computational Biology in Medicine, Weill Cornell Medical College, 1300 York Avenue, Box 194, New York, NY 10065, United States
[b] Feil Family Brain and Mind Research Institute and Department of Neurology, Weill Cornell Medical College, 1300 York Avenue, New York, NY 10021, United States

HIGHLIGHTS

- We present a method for power spectral estimation based on robust statistics.
- Compared to standard methods, the new approach is resistant to transient artifacts.
- Confidence intervals estimated in a Bayesian fashion have appropriate coverage.
- The approach is computationally efficient.
- Software is provided in the form of a MATLAB toolbox.

ARTICLE INFO

Article history:
Received 15 October 2015
Received in revised form 17 March 2016
Accepted 16 April 2016
Available online 19 April 2016

Keywords:
Spectral analysis
Fourier decomposition
Multitaper method
Bayesian confidence interval estimation
Order statistics

ABSTRACT

Background: Typical electroencephalogram (EEG) recordings often contain substantial artifact. These artifacts, often large and intermittent, can interfere with quantification of the EEG via its power spectrum. To reduce the impact of artifact, EEG records are typically cleaned by a preprocessing stage that removes individual segments or components of the recording. However, such preprocessing can introduce bias, discard available signal, and be labor-intensive. With this motivation, we present a method that uses robust statistics to reduce dependence on preprocessing by minimizing the effect of large intermittent outliers on the spectral estimates.
New method: Using the multitaper method (Thomson, 1982) as a starting point, we replaced the final step of the standard power spectrum calculation with a quantile-based estimator, and the Jackknife approach to confidence intervals with a Bayesian approach. The method is implemented in provided MATLAB modules, which extend the widely used Chronux toolbox.
Results: Using both simulated and human data, we show that in the presence of large intermittent outliers, the robust method produces improved estimates of the power spectrum, and that the Bayesian confidence intervals yield close-to-veridical coverage factors.
Comparison to existing method: The robust method, as compared to the standard method, is less affected by artifact: inclusion of outliers produces fewer changes in the shape of the power spectrum as well as in the coverage factor.
Conclusion: In the presence of large intermittent outliers, the robust method can reduce dependence on data preprocessing as compared to standard methods of spectral estimation.

1. Introduction

Electroencephalography (EEG), a technique for recording the electrical activity of the brain via surface electrodes, is a commonly used assay of brain activity in research and clinical settings. Well-recognized advantages of the EEG include its high temporal resolution, noninvasive nature, and ease of use (Bunge and Kahn, 2009). However, it is also highly sensitive to electrical activity from non-neural sources, such as eye movements (Gasser et al., 1992),

Abbreviations: EEG, electroencephalogram; PDF, probability density function; CDF, cumulative density function.

* Corresponding author.
E-mail addresses: tam2022@med.cornell.edu (T. Melman), jdvicto@med.cornell.edu (J.D. Victor).

http://dx.doi.org/10.1016/j.jneumeth.2016.04.015
0165-0270/

muscle activity (Whitham et al., 2007), electrode movement, and electric fields from the environment (Tatum et al., 2011). These sources generate signals that corrupt the underlying neural signal, and are difficult, if not impossible, to avoid.

For many research applications, and increasingly for clinical applications (Schiff et al., 2014), spectral measures are used to analyze EEG characteristics (Mitra and Pesaran, 1999). Since activity in specific frequency bands often has direct biological interpretations (Penfield and Jasper, 1954), the power spectrum is of particular interest. However, since the raw EEG signal is contaminated by non-neural sources, obtaining reliable estimates of the power spectrum that reflect underlying brain activity is not straightforward.

Computation of the power spectrum typically involves segmenting the continuous signal, applying Fourier analysis to each segment, and calculating the mean over segments of the power at each frequency. The data segments, typically of duration 1 s or more, may be determined arbitrarily (e.g., for records of spontaneous EEG), or based on events in a behavioral paradigm (e.g., for event-related potential studies). Fourier components arising from segments contaminated by typical artifacts (e.g., muscle and eye movements) are typically large relative to those of segments that only contain the neural signal, and therefore bias the mean upwards. This problem is usually addressed by removing these artifacts, by a combination of manual identification of artifact-containing segments and automated means, such as independent component analysis (ICA) (Makeig et al., 1996); however this can be labor- and time-intensive, subjective, and can discard portions of usable data.

Here we describe an alternative approach to this outlier problem, via the use of robust statistics. Specifically, we focus on the median and other quantile-based statistics. Via simulations and application to real EEG data, we show that this approach can recover the power spectrum of the underlying signal even in the presence of substantial artifact. Finally, we provide code that extends the Chronux (Bokil et al., 2010; Mitra and Bokil, 2008), toolbox to carry out these computations, including the calculation of Bayesian confidence intervals.

2. Methods

2.1. Algorithm

2.1.1. Modified multitaper method

A power spectrum is typically estimated from a measured time series by cutting the time series into segments, applying Fourier analysis to these segments, and averaging the power in each frequency bin across segments. The true value of the power spectrum is the limit of this process as the length and number of the data segments tend to infinity. However, in practice these segments are finite in length and limited in number, so power spectral estimates are necessarily biased (resulting from spectral leakage due to the finite length of the data segment) and imprecise (due to the finite number of data segments).

The multitaper method (Prieto et al., 2007), a power-spectral estimator that we use as a starting point for our approach, tackles the tradeoff between this bias and variance in a way that is optimal for Gaussian signals. The method minimizes spectral leakage (the artifactual spreading of power from one frequency bin into its neighbors), by windowing each segment by an orthogonal set of functions, the Slepian tapers. For further background on the multitaper method see Thomson (1982), Mitra and Pesaran (1999) and Mitra and Bokil (2008). Chronux is a freely available MATLAB toolbox that provides convenient implementations of the multitaper method, which we extend with an implementation of the robust approach.

The standard multitaper method consists of the following steps: (1) multiplying each data segment by each of the tapers, (2) applying Fourier analysis to these products, (3) averaging over tapers within each segment, and (4) averaging over segments. To formalize this, we denote the original signal by $X(t)$, with B segments cut from the signal denoted as $x_1(t), \ldots, x_b(t), \ldots, x_B(t)$, each of length T. These segments are non-overlapping, but need not be contiguous. We denote the K Slepian tapers by $a_1(t), \ldots, a_k(t), \ldots, a_K(t)$ (the choice of K is driven by the desired spectral resolution and data length; a common choice for 3-s-long segments, and the Chronux default, is $K = 5$). With this notation, the standard multitaper estimate of $S_x(\omega)$, the true spectral power at frequency ω, is defined as:

$$\hat{S}^{\text{standard}}(\omega) = \frac{1}{B}\sum_{b=1}^{B}\sum_{k=1}^{K}\frac{1}{T}\left|\int_0^T x_b(t)\,a_k(t)\,e^{-i\omega t}dt\right|^2. \quad (1)$$

We denote the power estimate for a single sample b and a single taper by $S_{b,k}(\omega)$:

$$S_{b,k}(\omega) = \frac{1}{T}\left|\int_0^T x_b(t)\,a_k(t)\,e^{-i\omega t}dt\right|^2. \quad (2)$$

With this notation, the standard spectral estimate takes the form

$$\hat{S}^{\text{standard}}(\omega) = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{K}\sum_{k=1}^{K}S_{b,k}(\omega). \quad (3)$$

Thus, the standard multitaper estimate is a nested mean: first a mean over the K tapers within each segment to obtain the estimate $\hat{S}_b(\omega) = \text{mean}\left(\left\{\hat{S}_{b,k}(\omega)\right\}\right)$, and then a mean over the B segments:

$$\hat{S}^{\text{standard}}(\omega) = \text{mean}\left(\left\{\hat{S}_b(\omega)\right\}\right). \quad (4)$$

Since our goal is to reduce the effect of outlier estimates from each segment, we replace the mean over segments by a robust estimator, resulting in the estimated power spectral quantity $\hat{S}^{\text{robust}}(\omega)$. There are many possible choices for the robust estimator—for example: an estimator based on the h^{th} quantile, a trimmed mean, a Winsorized mean (Huber, 1963), or iterative rejection of outliers. While the present framework applies to all of these choices, estimators based on quantiles are more amenable to computation of Bayesian confidence intervals (see below), and we therefore focus on these, both in the illustrations below and in the MATLAB toolbox. We denote the estimator based on the h^{th} quantile as $\hat{S}^{\text{quantile\_}h}(\omega)$. Note that $h = 1/2$ corresponds to the median; this is the default value in the code.

Even for Gaussian data, the median power of the tapered estimates does not equal the mean power. This is because spectral estimates are approximately distributed as chi-squared, which is positively skewed. As shown in Appendix A, we can take the skewing into account by dividing the median power by a data-independent scale factor. Furthermore, scale factors can be derived that convert not just the median (0.5 quantile), but any quantile, into mean power. Appendix A details the calculation of these scale factors, which is implemented in the MATLAB module analytical_scalefactor_Robust().

Including this scale factor yields our main result, the robust spectral estimate:

$$\hat{S}^{\text{quantile\_}h}(\omega) = \frac{\text{quantile\_}h\left(\left\{\hat{S}_b(\omega)\right\}\right)}{C(h,d,B)}, \quad (5)$$

where $C(h, d, B)$ is the scale factor for quantile h; d is the number of degrees of freedom ($d = 2K$ for typical frequencies, $d = K$ for DC and the Nyquist frequency); and B, as above, is the number of segments.