A Review of Multitaper Spectral Analysis

Behtash Babadi*, *Member, IEEE*, and Emery N. Brown, *Fellow, IEEE*

***Abstract*—Nonparametric spectral estimation is a widely used technique in many applications ranging from radar and seismic data analysis to electroencephalography (EEG) and speech processing. Among the techniques that are used to estimate the spectral representation of a system based on finite observations, multitaper spectral estimation has many important optimality properties, but is not as widely used as it possibly could be. We give a brief overview of the standard nonparametric spectral estimation theory and the multitaper spectral estimation, and give two examples from EEG analyses of anesthesia and sleep.**

***Index Terms*—Electroencephalogram (EEG) analysis, multitaper, spectral estimation.**

I. Introduction

MANY natural systems exhibit oscillatory dynamics. Examples include electroencephalogram (EEG) and local field potential recordings from the human brain [1], speech [2], oceanography [3], climatic time series [4], and seismic data [5]. Characterizing such oscillatory behavior gives insight into the underlying dynamics of these systems. However, it is often not clear how to efficiently analyze the output of the system in the *time* domain. A *frequency* domain or spectral analysis of the output is often more informative. In order to conduct a spectral analysis, we need to address two main issues. First, in order to describe a system in frequency domain, in principle, we need to have a sample of infinite length of the output of the system. Second, most of the observable natural phenomena exhibit stochasticity, which can be due to the intrinsic properties of the system or can be an artifact of the observation. Therefore, infinitely many realizations of the output of the system are required to capture the stochastic properties. However, in most problems, we can only observe the output of these systems as a single realization with finite length.

The theory of spectral estimation aims at addressing these issues, and thus provides a framework to estimate the frequency-domain representation of the output of a stochastic system based on finite data. Finite data typically result in spectral estimates which are both biased and suffer from high error variance. Several methods have been suggested in the nonparametric spectral estimation literature to reduce the bias (e.g., tapering) or the variance (e.g., lag window smoothing) of the estimates [6], but they usually depend on *ad hoc* parameters which are often challenging to tune in practice.

Arguably, the first nonparametric method to address the issues of bias and variance simultaneously in an "optimal" fashion appeared with the introduction of the multitaper spectral estimation technique [7]. Multitaper spectral analysis has gained popularity among some researchers in bioengineering and neuroscience in recent years (see, for example, [8]–[13]). However, it is not as widely used as it possibly could be. Although the original presentation by Thomson is elegant and rigorous, it is not readily accessible to the general audience. We review the conceptual framework of multitapering for the broader audience in biomedical engineering. We give a brief overview of the bias and variance problems by examining two of the basic nonparametric spectral estimation techniques, namely the periodogram and tapered spectral estimates. We describe how the multitaper spectral estimation simultaneously addresses the issues of bias and variance. Finally, we present two applications of multitaper spectral estimation in EEG analyses in anesthesia and sleep to highlight the significance of this technique in real data applications.

II. Second-Order Stationary Random Processes and Their Spectral Representation

Let $x(t)$ be the output of a stochastic system at time t and suppose that the output is observed at a sampling rate of $F_s := \frac{1}{\Delta}$, where Δ denotes the sampling interval. We use the symbol ":=" to indicate definitions. The equivalent observed signal in discrete time can be defined as $x_k := x(k\Delta)$, for $k = 0, 1, 2, \ldots$. The most intuitive approach to developing a spectral representation of a stochastic system is to compute the Fourier transform of its output. However, the Fourier transform of a random signal is in general undefined, due to lack of integrability in continuous time or summability in discrete time [6]. Let us instead consider a sample of the output of size N, given by x_k for $k = 0, 1, \ldots, N-1$. The Fourier transform of x_k can be written as

$$X(f) := \Delta \sum_{k=0}^{N-1} x_k e^{-i2\pi k f \Delta}. \tag{1}$$

The power of the signal x_k in the band $f \pm \frac{1}{2}\delta f$ is approximately given by

$$P := \frac{1}{N\Delta} |X(f)|^2 \delta f. \tag{2}$$

Manuscript received October 10, 2013; revised February 2, 2014 and March 5, 2014; accepted March 7, 2014. Date of publication March 14, 2014; date of current version April 17, 2014. This work was supported by: NIH Award R01GM10498 (ENB). *Asterisk indicates corresponding author.*

*B. Babadi is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: behtash@umd.edu).

E. N. Brown is with the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139 USA, with the Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Boston, MA 02114 USA, with the Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA 02139 USA, and with the Institute for Medical Engineering and Science, Cambridge, MA 02139 USA (e-mail: enb@neurostat.mit.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2014.2311996

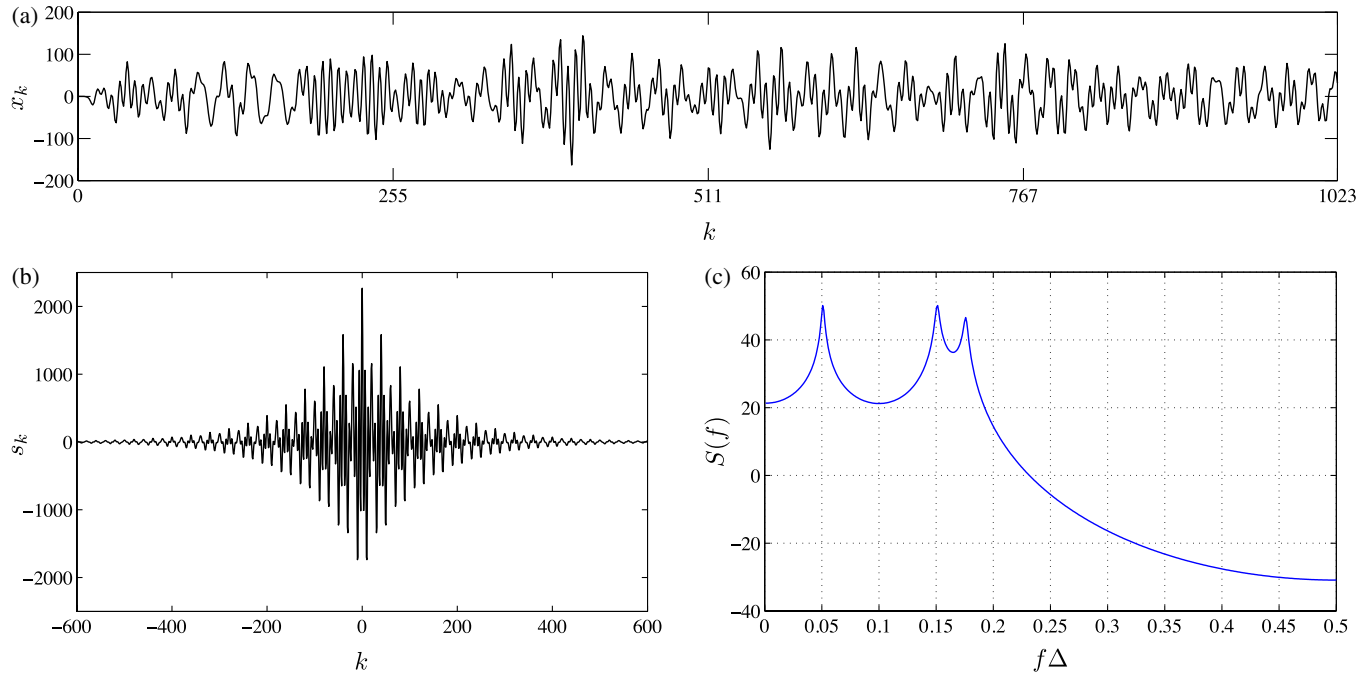

Fig. 1. The AR(6) process given by (7). (a) Sample of length 1024, (b) the autorcovariance sequence, and (c) the PSD.

If we take the expectation of the power and let $N \to \infty$, the quantity

$$\lim_{N\to\infty} \mathbb{E}\left\{\frac{1}{N\Delta}|X(f)|^2\right\}\delta f \tag{3}$$

gives the average power of the signal around frequency f, where $\mathbb{E}\{\cdot\}$ is the ensemble expectation. But, this quantity is impossible to compute in practice, since it requires infinitely many samples ($N \to \infty$) and infinitely many realizations of the output signal. Therefore, the standard approach in spectral analysis is to restrict analyses to the class of random signals which are second-order stationary (s.o.s.) and *ergodic*. For an s.o.s. process x_k, $k = 0, 1, 2, \ldots$, the *mean* is independent of k:

$$\mu_i := \mathbb{E}\{x_i\} = \mu, \quad i =, 0, 1, \ldots \tag{4}$$

and the *autocovariance* sequence satisfies

$$s_{i,j} := \mathbb{E}\big\{(x_i - \mu_i)(x_j - \mu_j)\big\} = s_{|i-j|}, \quad i, j = 0, 1, \ldots \tag{5}$$

Hence, the autocovariance sequence can be expressed as s_k, for $k = 0, 1, \ldots$. The process is *ergodic*, if the ensemble expectation equals the time expectation (see [14] for an extended discussion).

Examples of ergodic s.o.s. signals are the white noise process, autoregressive processes, and the harmonic process. In particular, s.o.s. random processes admit a spectral representation by virtue of the Wiener–Khinchin theorem [15]. This theorem states that for discrete s.o.s. stochastic processes,

$$\lim_{N\to\infty} \mathbb{E}\left\{\frac{1}{N\Delta}|X_p(f)|^2\right\} =: S(f) = \Delta \sum_{k=-\infty}^{\infty} s_k e^{-i2\pi k f \Delta}. \tag{6}$$

Comparing (3) and (6), we observe that $S(f)$ is the power density of the random signal x_k at frequency f and is thus termed the *power spectral density* (PSD). The Wiener–Khinchin theorem shows that for s.o.s. random signals, the Fourier transform of the autocovariance sequence is equivalent to the PSD.

If we know the autocovariance sequence s_k of an ergoidic s.o.s. signal x_k for $-\infty < k < \infty$, then, in principle, the spectral density can be computed. But, computing the autocovariance requires carrying out time averages over an infinite sample size of the random signal x_k. The main question of spectral estimation for s.o.s. random signals is: how can we estimate s_k, given a sample of finite length of a random signal x_k, so that its Fourier transform is a reliable estimate of the PSD? Let $\widehat{S}(f)$ be an estimate of the true PSD $S(f)$ of a random signal x_k. We require that a reliable estimator of $S(f)$ satisfies:

1) Approximate unbiasedness: $\mathbb{E}\big\{\widehat{S}(f)\big\} \approx S(f)$.

2) Low variance: $\mathsf{Var}\big\{\widehat{S}(f)\big\} \approx 0$.

As an illustration of an ergodic s.o.s. process with known properties, we use a sixth-order autoregressive process, denoted henceforth by the AR(6) process, given by

$$\begin{aligned} x_k = {} & 3.9515x_{k-1} - 7.8885x_{k-2} + 9.7340x_{k-3} \\ & - 7.7435x_{k-4} + 3.8078x_{k-5} - 0.9472x_{k-6} + v_k \end{aligned} \tag{7}$$

where v_k is zero-mean independent, identically distributed Gaussian noise with unit variance. Fig. 1(a) shows a sample of length 1024 from the AR(6) process, which is used throughout this study for spectral estimation. The autocovariance sequence and the PSD of the AR(6) process are shown in Fig. 1(b) and (c), respectively.

III. Periodogram and Tapering

A. The Periodogram

Suppose that a sample of size N of the random signal x_k is given. The most intuitive approach to obtain a nonparametric estimate of the autocovariance sequence s_k is the method of moments, where the estimate is based on the sample mean and sample autocovariance of the data

$$\widehat{\mu} = \frac{1}{N}\sum_{i=0}^{N-1} x_i \tag{8}$$

and

$$\widehat{s}_k = \frac{1}{N}\sum_{i=0}^{N-k-1}(x_i - \widehat{\mu})(x_{i+k} - \widehat{\mu}). \tag{9}$$

Taking the Fourier transform of the above estimate gives

$$\widehat{S}(f) = \Delta\sum_{k=-N+1}^{N-1}\widehat{s}_k e^{-i2\pi k f\Delta} = \frac{\Delta}{N}\left|\sum_{k=0}^{N-1} x_k e^{-i2\pi k f\Delta}\right|^2. \tag{10}$$

The estimate of (10) is called the periodogram and is attributed to Schuster [16]. Fig. 2(a) shows the periodogram estimate of the PSD of the AR(6) process. The periodogram estimate has two drawbacks that are evident in Fig. 2(a). First, beyond $f\Delta \approx 0.25$, the estimate has a significant deviation from the true PSD. Moreover, at lower frequencies, the estimate has a very high variability. Indeed, we can compute the expectation of the periodogram estimate as [6]

$$\mathbb{E}\{\widehat{S}(f)\} = \int_{-1/2\Delta}^{1/2\Delta} N\Delta\mathcal{D}_N^2\big((f-f')\Delta\big)S(f')df' \tag{11}$$

where

$$\mathcal{D}_N^2(f\Delta) := \frac{\sin^2(N\pi f\Delta)}{N^2\sin^2(\pi f\Delta)}. \tag{12}$$

The function $N\Delta\mathcal{D}_N^2(f\Delta)$ is called the Fejér kernel. Equation (11) shows that the expectation of the periodogram estimate is a smoothed version of the true PSD, where the smoothing kernel is $N\Delta\mathcal{D}_N^2(f\Delta)$. Fig. 2(b) shows the Fejér kernel centered at $f\Delta = 0.4$. The deviation of the expectation of the estimate from the true PSD is termed bias. The bias at a given frequency can be roughly decomposed into a narrow-band bias (resolution) and a broad-band bias [6]. The narrow-band bias comes from the main lobe of the smoothing kernel, which blurs the estimate within the width of the main lobe. The broad-band bias results from the contribution of the side lobes of the smoothing kernel, which leaks the broad components of the PSD into the estimate at any given frequency. For instance, at $f\Delta = 0.4$, the broad-band bias has a significant contribution, since the side lobes are aligned with the peaks of the true PSD.

B. Bias Reduction via Tapering

Tapering, originally proposed by Blackman and Tukey [17], is an effective way to tradeoff the broad-band and narrow-band bias of spectral estimates. Consider a taper h_k, $k = 0, 1, 2, \ldots, N-1$, which is multiplied by the data x_k to form the so-called tapered data. The peridogram of the tapered data constitutes the tapered spectral estimate:

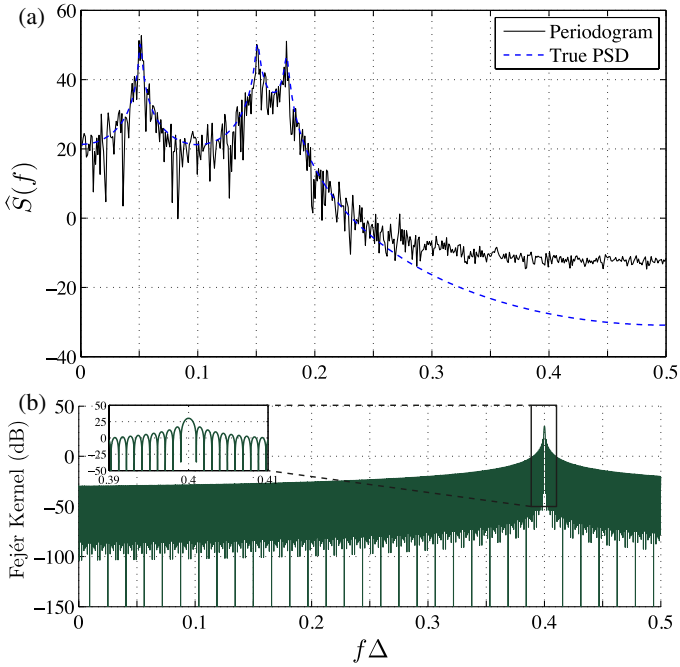

Fig. 2. Periodogram estimate of the PSD of the AR(6) process. (a) Periodogram estimate. (b) Fejér kernel centered at $f\Delta = 0.4$. The zoomed-in view of the main lobe is shown on the top left.

$$\widehat{S}^{\mathrm{t}}(f) = \Delta\left|\sum_{k=0}^{N-1} h_k x_k e^{-i2\pi k f\Delta}\right|^2. \tag{13}$$

The expectation of the tapered estimate is given by [6]

$$\mathbb{E}\{\widehat{S}^t(f)\} = \int_{-1/2\Delta}^{1/2\Delta}\frac{1}{\Delta}\big|H(f-f')\big|^2 S(f')df' \tag{14}$$

where

$$\big|H(f)\big|^2 = \Delta^2\left|\sum_{k=0}^{N-1} h_k e^{-i2\pi k f\Delta}\right|^2. \tag{15}$$

Equation (14) implies that the expectation of the tapered estimate is a smoothed version of the true PSD, where the smoothing kernel is given by $\frac{1}{\Delta}|H(f)|^2$. Therefore, by designing the taper h_k with small side lobes, the broad-band bias can be reduced. Many tapers have been proposed in the spectral estimation literature [6]. As an example, Fig. 3(a) shows the tapered estimate of the AR(6) process using the Hann taper given by

$$h_k = \frac{1}{2}\left(1 - \cos\left(\frac{2\pi k}{N-1}\right)\right), \quad k = 0, 1, \ldots, N-1. \tag{16}$$

Fig. 3(a) implies that the broad-band bias is significantly reduced. Fig. 3(b) shows the smoothing kernel $\frac{1}{\Delta}|H(f)|^2$ corresponding to the Hann taper centered at $f\Delta = 0.4$. The side lobes of the Hann taper are much smaller than those of the Fejér kernel, and hence, the broad-band leakage is negligible. However, the main lobe of the Hann taper is wider than that of the

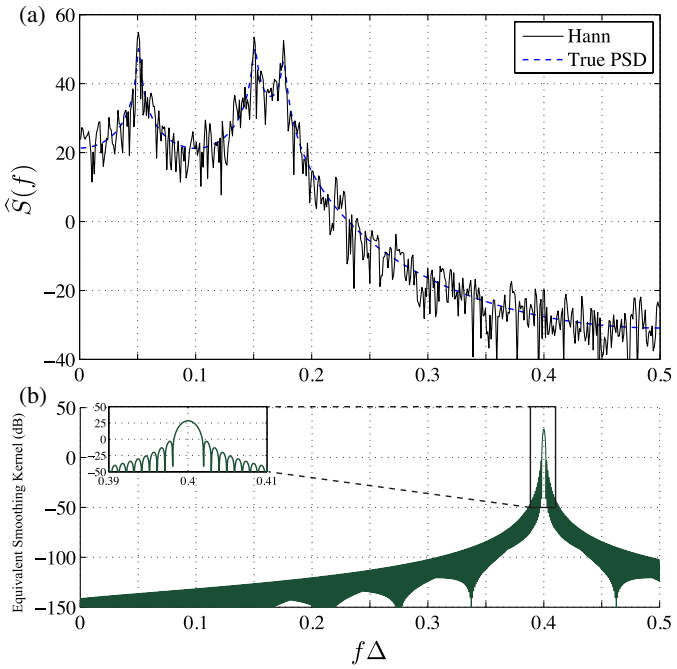

Fig. 3. Tapered estimate of the PSD of the AR(6) process. (a) Tapered estimate using the Hann taper. (b) Equivalent smoothing kernel $\frac{1}{\Delta}|H(f)|^2$ centered at $f\Delta = 0.4$. The zoomed-in view of the main lobe is shown on the top left.

Fejér kernel. As a rule of thumb, the wider the main lobe, the smaller the side lobes would be [6]. Therefore, tapering can be interpreted as a tradeoff between the narrow-band bias (resolution) and the broad-band bias. A natural question that arises is which tapers are "good"? or, which taper is the "optimal" taper? As we will see in the next section, multitaper spectral estimation addresses this issue.

C. Inconsistency of Tapered Spectral Estimates

The second drawback of the periodogram estimate is its high variance. Figs. 2(a) and 3(a) show that the periodogram and tapered estimates have a high variance. Indeed, the variance of both estimates can be computed as [6]

$$\mathsf{Var}\{\widehat{S}^{\mathsf{t}}(f)\} \overset{N\gg 1}{\approx} \begin{cases} 2S^2(f), & \text{for} f \neq 0 \text{ and } 1/2\Delta \\ S^2(f), & \text{otherwise.} \end{cases} \tag{17}$$

The expression for the variance implies that both estimators are inconsistent, i.e., as $N \to \infty$, the variance would not tend to zero, or equivalently, the spectral estimate would not converge to the true PSD. As we will see in the next section, multitaper spectral estimation addresses this issue as well.

IV. Multitaper Spectral Estimation

Multitaper spectral estimation theory was developed through the seminal work of David J. Thomson in 1982 [7]. The goal of this theory is to address the issues of bias and variance of nonparametric spectral estimation simultaneously in an "optimal" fashion. Since then, several variants of the multitaper spectral estimator have appeared in the literature (see, for example, [18]). We give an overview the original multitaper spectral estimation technique.

Before going into the details of multitaper spectral estimation, we begin with a simple example for motivation. Suppose that we want to estimate the mean μ of a Gaussian distribution with unit variance by observing one sample, z_1. The maximum-likelihood (ML) estimate of the mean is $\widehat{\mu} = z_1$, and the variance of the estimate is $\mathsf{Var}(\widehat{\mu}) = 1$. Next, consider L *uncorrelated* samples of the same distribution, given by $z_1, z_2, \ldots, z_L$. The ML estimate of the mean is now given by $\widehat{\mu} = \frac{1}{L}\sum_{i=1}^{L} z_i$, and the variance is given by $\mathsf{Var}(\widehat{\mu}) = 1/L$. In both cases, the estimates are unbiased, but in the latter, the variance is reduced by a factor of L. This simple example is the main idea behind multitapering: *given L "good" and "uncorrelated" tapers, the multitaper spectral estimate formed by averaging the corresponding L tapered estimates would have a variance reduced by a factor of L.* This idea was first proposed by Bartlett [19] and later by Welch [20], but was carried out in a principled way by Thomson [7]. We now state this more formally.

A. Derivation of the Multitaper Spectral Estimate

Let $h_k^{(1)}, h_k^{(2)}, \ldots, h_k^{(L)}$ be a set of tapers. The multitaper spectral estimate is given by

$$\widehat{S}^{\mathsf{mt}}(f) := \frac{1}{L}\sum_{i=1}^{L} \widehat{S}^{(i)}(f) \tag{18}$$

where

$$\widehat{S}^{(i)}(f) := \Delta \left| \sum_{k=0}^{N-1} h_k^{(i)} x_k e^{-i2\pi k f \Delta} \right|^2. \tag{19}$$

If the estimates $\widehat{S}^{(i)}(f)$ are *uncorrelated*, then the variance of the multitaper estimate would be $\approx \frac{S^2(f)}{L}$. To guarantee approximate uncorrelatedness of the tapered estimates $\widehat{S}^{(i)}(f)$, the tapers must be orthonormal [6]:

$$\sum_{k=0}^{N-1} h_k^{(i)} h_k^{(j)} = \delta_{ij}. \tag{20}$$

Therefore, the problem reduces to finding a set of L orthonormal tapers to minimize the bias of the estimate. It is not hard to show that the expectation of the multitaper estimate is given by [6]

$$\mathbb{E}\{\widehat{S}^{\mathsf{mt}}(f)\} = \frac{1}{L}\sum_{j=1}^{L} \int_{-1/2\Delta}^{1/2\Delta} \frac{1}{\Delta} \left|H^{(j)}(f-f')\right|^2 S(f')df' \tag{21}$$

where $H^{(j)}(f)$ is the Fourier transform of $h_k^{(j)}$. Let

$$R := \frac{2\alpha}{N\Delta} \tag{22}$$

be the designed spectral resolution, for some constant $\alpha \geq 1$. The constant α is denoted by the time-bandwidth product. If N is large enough so that R is small, the expectation can be

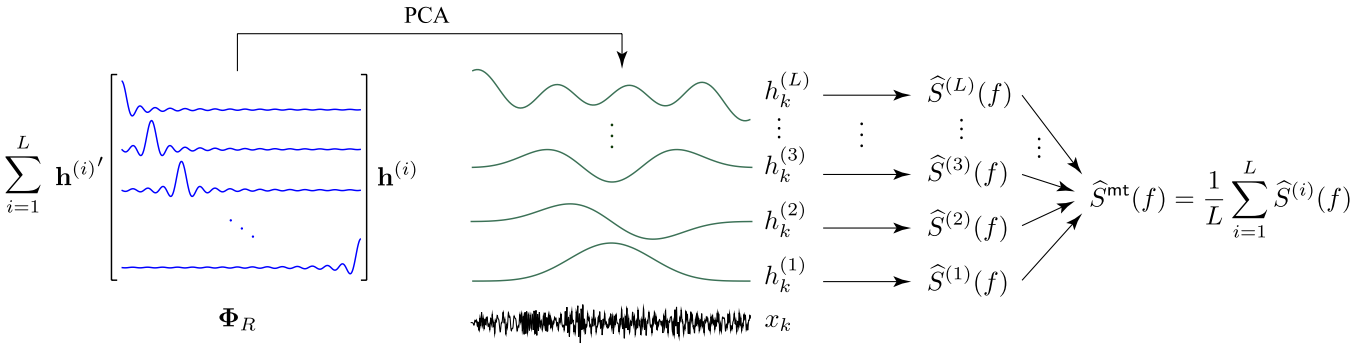

Fig. 4. Schematic depiction of multitaper spectral estimation.

upper-bounded as [21]:

$$\mathbb{E}\{\widehat{S}^{\mathsf{mt}}(f)\} \leq \max_f S(f)\left(1-\frac{1}{L}\sum_{j=1}^{L}\frac{1}{\Delta}\int_{-R/2}^{R/2}|H^{(j)}(f')|^2 df'\right) + S(f) + \mathcal{O}(R). \tag{23}$$

Instead of minimizing the bias itself, we seek to minimize the above upper bound [21]. That is, we seek a set of L tapers that maximize

$$\frac{1}{L}\sum_{i=1}^{L}\int_{-R/2}^{R/2}\left|H^{(i)}(f')\right|^2 df'. \tag{24}$$

The above maximization problem can be recast in the time domain as follows [6]:

$$\max_{h_k^{(1)},h_k^{(2)},\cdots,h_k^{(L)}} \frac{1}{L}\sum_{j=1}^{L}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1} h_k^{(j)}\frac{\sin\left(\pi R\Delta(k-l)\right)}{\pi(k-l)}h_l^{(j)}. \tag{25}$$

Let $\mathbf{h}^{(i)} := [h_0^{(i)}, h_1^{(i)}, \ldots, h_{N-1}^{(i)}]'$, and $\mathbf{\Phi}_R$ be an $N \times N$ symmetric matrix with elements $(\mathbf{\Phi}_R)_{k,l} := \frac{\sin\left(\pi R\Delta(k-l)\right)}{\pi(k-l)}$. Then, the above maximization problem can be expressed as

$$\max_{\mathbf{h}^{(1)},\mathbf{h}^{(2)},\cdots,\mathbf{h}^{(L)}} \sum_{i=1}^{L}\mathbf{h}^{(i)\prime}\mathbf{\Phi}_R\mathbf{h}^{(i)}. \tag{26}$$

The above equation resembles the principal component analysis (PCA) of the matrix $\mathbf{\Phi}_R$, and the solution is indeed given by the L eigenvectors of the matrix $\mathbf{\Phi}_R$ corresponding to its L highest eigenvalues. Fig. 4 shows a schematic depiction of the multitaper spectral estimation. Given R, one can find these eigenvectors by performing PCA on $\mathbf{\Phi}_R$ and picking the first L eigenvectors corresponding to the L highest eigenvalues. These eigenvectors are known as discrete prolate spheroidal sequences (dpss) or Slepian sequences [6], [22]. They form a set of orthonormal tapers, which have the highest possible concentration of energy in the band $[-R/2, R/2]$. In order for the L tapers to have desirable concentration properties in the band $[-R/2, R/2]$, it can be shown that [6] a good choice of L is

$$L \ll \lfloor 2\alpha \rfloor - 1. \tag{27}$$

Fig. 5 shows some of the dpss for $N = 128$ and $\alpha = 4$. The above bound in this case implies that L must be significantly less than 7. As Fig. 5 shows, the first and second dpss sequences have a very high concentration in the band $[-\frac{1}{32\Delta}, \frac{1}{32\Delta}]$, but the seventh sequence exhibits a high degree of broad-band leakage.

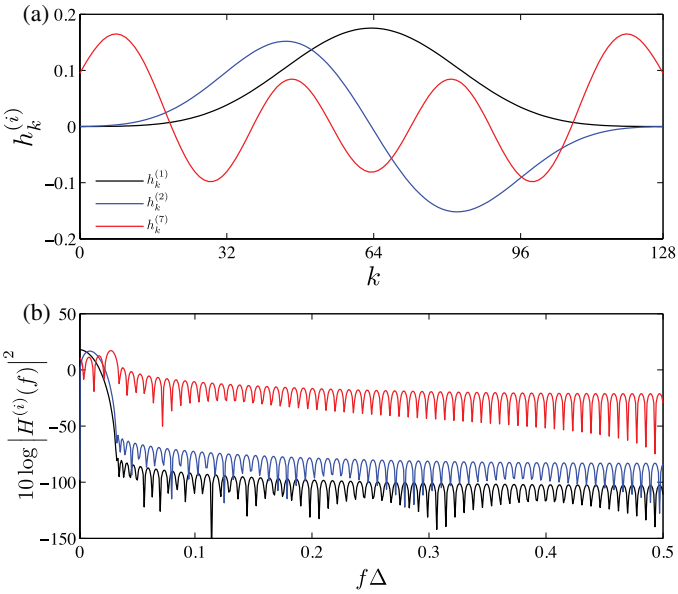

Fig. 5. Example of dpss for $N = 128$ and $R = 8/N\Delta$. (a) the first (black), second (blue) and seventh (red) dpss, and (b) the equivalent smoothing kernels.

Since for $L \ll \lfloor 2\alpha \rfloor - 1$ the dpss sequences have very high concentration of energy in the main lobe, the narrow-band bias or the spectral resolution can be very well approximated by the width of the main lobe, which is the designed spectral resolution R. Fixing the spectral resolution identifies α, which can be thought of as the initial step of the multitaper spectral estimation. Given data of length N and sampling interval Δ, the multitaper spectral estimation can be summarized as follows:

1) Fix the spectral resolution R, such that $\alpha = \frac{N\Delta R}{2} \geq 1$.
2) Fix $L \ll \lfloor 2\alpha \rfloor - 1$.
3) Form the matrix $\mathbf{\Phi}_R$, and find its L eigenvectors corresponding to the L highest eigenvalues.
4) Form the individual tapered spectral estimates using the L dpss sequences, and average them.

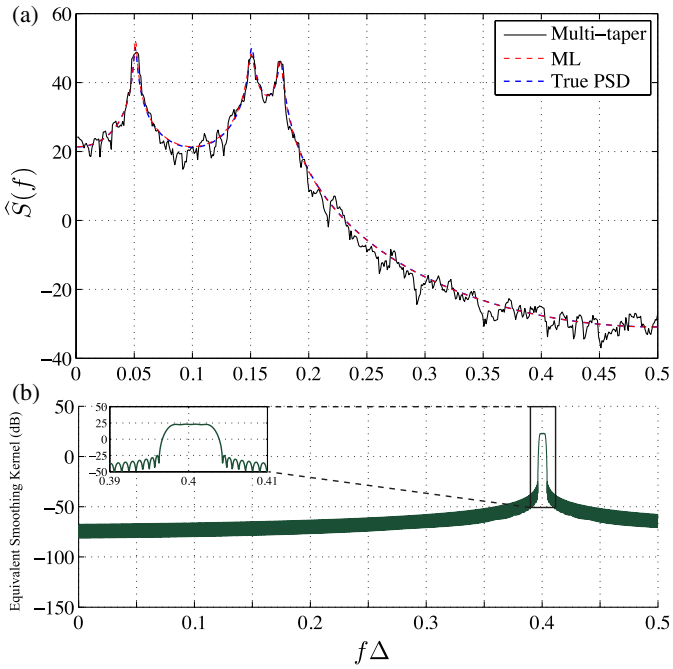

Fig. 6. Multitaper and ML estimates of the PSD of the AR(6) process with $R = 8/N\Delta$ and $L = 4$. (a) the multitaper and ML estimates, (b) the equivalent smoothing kernel centered at $f\Delta = 0.4$. The zoomed-in view of the main lobe is shown on the top left.

Note that in practice, step 3 can be carried out using standard software which generate the dpss sequences given R and L.

For the AR(6) process, we fix a spectral resolution of $R = \frac{1}{128\Delta}$, which implies $\alpha = 4$. We thus use $L = 4$ tapers to form the multitaper estimate. Fig. 6(a) shows the multitaper estimate of the PSD of the AR(6) process. Comparing this estimate to those of the periodogram (see Fig. 2) and the Hann taper (see Fig. 3), we see that both the bias and variance are appreciably reduced. In Fig. 6(a), we also show the approximate ML estimate of the AR(6) process using the least squares method [6]. The ML estimate is a parametric estimate, which assumes that the data are generated by an AR model and directly estimates the AR parameters. The ML estimate as a parametric method is more robust to finite data than nonparametric methods [6]. In general, when a parametric model assumption is valid for the data, parametric methods are more accurate and more efficient than nonparameteric methods [6]. However, parametric methods require a model order selection step [23] analogous to the choice of number of tapers for the multitaper method. We refer the readers to [6] for a detailed discussion of parametric spectral estimation. Moreover, as it will be discussed later in Section V, nonparametric methods such as the multitaper estimate are more suitable when dealing with nonstationary data and their spectrogram representation.

Fig. 6(b) shows the equivalent smoothing kernel of the four dpss tapers, which forms the bias. As shown in Fig. 6(b), the smoothing kernel has a very high side lobe suppression which results in a significantly reduced broad-band leakage. In particular, the first side lobe is about 50 dB lower than the main lobe. As a result, the energy is highly concentrated within the main lobe and, as expected by design, the spectral resolution can be very well approximated by the width of the main lobe $R = \frac{1}{128\Delta}$. Note that the main lobe of the equivalent smoothing kernel of the multitaper estimate in Fig. 6(b) is wider than that of the periodogram [see Fig. 2(b)] and the Hann tapered estimate [see Fig. 3(b)] by design. A wider main lobe results from choosing a higher value of the time-bandwidth product α, which in turn admits the use of more tapers with good concentration properties. This later fact can be viewed as the tradeoff between the narrow-band bias and variance.

The multitaper estimate is approximately unbiased:

$$\mathbb{E}\{\widehat{S}^{\mathsf{mt}}(f)\} \approx S(f) \tag{28}$$

if only significantly less than $\lfloor 2\alpha \rfloor - 1$ tapers are used [21]. Also, if the true PSD is uniformly continuous, the variance can be upper-bounded as [21]

$$\mathsf{Var}\{\widehat{S}^{mt}(f)\} \leq \begin{cases} 2S^2(f)/L, & \text{for} f \neq 0 \text{ and } 1/2\Delta \\ S^2(f)/L & \text{otherwise} \end{cases} + \mathcal{O}\left(\frac{1}{L}\right) + \mathcal{O}(R). \tag{29}$$

The second and thirds terms are negligible for large enough L and small enough R. Comparing (29) with (17), the variance of the multitaper estimate is reduced by a factor of L. Since $\alpha = RN\Delta/2$, for a fixed resolution R, as $N \to \infty$, we have $\alpha \to \infty$. Also, since $L \ll \lfloor 2\alpha \rfloor - 1$, one can choose L so that $L \to \infty$ [21]. Therefore, the multitaper estimate is asymptotically consistent, i.e., the variance goes to zero as $N \to \infty$.

It is possible to extend the multitaper method to multivariate time series [24]. Suppose that $\mathbf{x}_k := [x_{1,k}, x_{2,k}, \ldots, x_{p,k}]'$ is a p-dimensional s.o.s. time series. Then, the multitaper cross-spectral estimate of the lth and mth time series is given by [24]

$$\widehat{S}^{\mathsf{mt}}_{l,m}(f) := \frac{1}{L}\sum_{i=1}^{L} \widehat{S}^{(i)}_{l,m}(f) \tag{30}$$

where

$$\widehat{S}^{(i)}_{l,m}(f) := \Delta \left[\sum_{k=0}^{N-1} h_k^{(i)} x_{l,k} e^{-i2\pi k f \Delta}\right]\left[\sum_{k=0}^{N-1} h_k^{(i)} x_{m,k} e^{i2\pi k f \Delta}\right] \tag{31}$$

and L is defined the same as in the univariate case. The statistical properties of the multivariate multitaper spectral are discussed in detail in [24].

As mentioned earlier, other methods such as the Welch's method [20], address the problems of bias and variance in suboptimal ways. In the Welch's method, the data are divided into several segments in time domain, where for each segment a tapered estimate is computed. The spectral estimate is the average of the tapered estimates obtained from each segment of the data. The length of the data segments used in the Welch's methods limits the spectral resolution, but the averaging of tapered estimates from several segments reduces the variance of the final estimate. The multitaper method, however, provides a more favorable tradeoff between narrow-band bias, broad-band bias, and variance than the Welch's method [25].

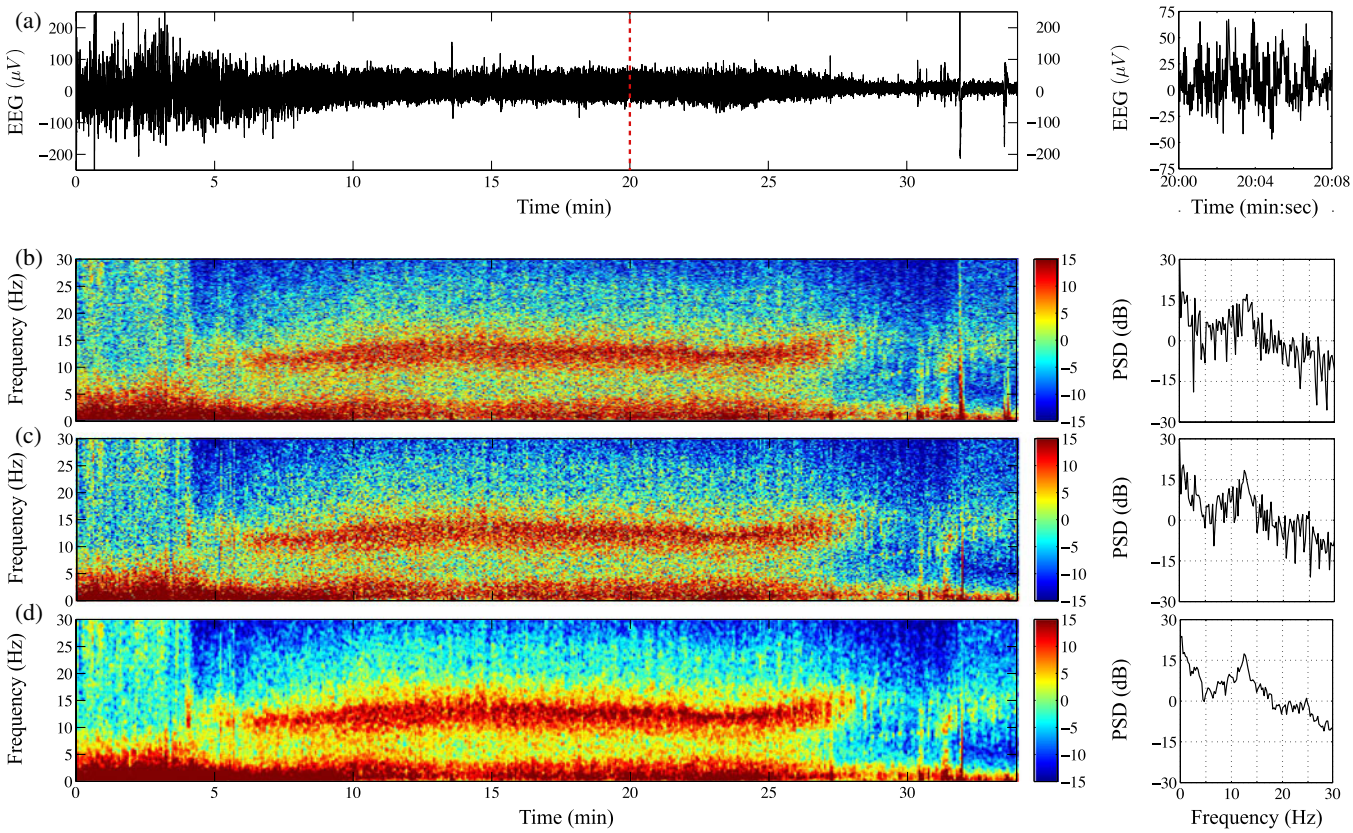

Fig. 7. Spectrogram of EEG during propofol-induced general anesthesia. (a) The time-domain EEG signal, (b) spectrogram from periodogram spectral estimates, (c) spectrogram from tapered spectral estimates using the Hann taper, and (d) spectrogram from multitaper spectral estimates. The right panels show the corresponding zoomed-in view of 8 s of data starting at minute 20 (marked with the red-dashed line in A).

V. Application to Real Data

Often in practice the assumption of second-order stationarity does not hold. Therefore, the spectral estimation techniques for s.o.s. processes cannot be readily applied to nonstationary signals. However, for a large class of nonstationary signals, such as EEG under sleep or anesthesia, the variability in the spectral characteristics of the signal occurs at a time scale much larger than the sampling interval. A widely used solution to spectral characterization of such nonstationary signals is the following: pick a window small enough so that the signal is assumed to be s.o.s., compute the nonparametric spectral estimate using techniques for s.o.s. processes, slide the window across time, and repeat. This idea leads to a spectral representation for nonstationary signals, termed the spectrogram representation.

The spectrogram is the concatenation of the PSD of the data in small sliding windows with overlap. The overlap makes it possible to compute spectral estimates at a desired temporal resolution. For each window of the observation, the spectral estimate is obtained using the estimation techniques common for s.o.s. processes. The choice of the window length determines the spectral resolution of the spectral estimate. However, the choice of the window length also limits the time resolution of the spectrogram, by the uncertainty principle. The larger the window length is chosen, the higher the spectral resolution and the lower the temporal resolution would be. More explicitly, consider a nonstationary signal x_n, $n = 0, 1, \ldots, M$. Computing the spectrogram representation based on multitaper spectral estimation can be summarized as follows:

1) Estimate the time-scale on which the data can be assumed to be s.o.s., and denote it by T. Let $N := \lfloor \frac{T}{\Delta} \rfloor$.
2) Pick a spectral resolution R, such that $\alpha = \frac{N\Delta R}{2} \geq 1$.
3) Pick a sliding length $\tau \leq T$, and let $K := \lfloor \frac{\tau}{T} \rfloor N$.
4) Find the multi-taper estimate of the PSD of the data at windows of length N, centered at time indices $n = \frac{N}{2}, \frac{N}{2} + K, \frac{N}{2} + 2K, \cdots$ until the entire data series is covered.

It is possible to construct the spectrogram based on a parametric spectral estimator. At each window, assuming an AR model, a parametric estimate of the spectral density can be obtained through model selection and fitting. However, when dealing with long datasets, model selection at each window results in a higher computational complexity than that required by the nonparametric spectrogram.

To illustrate the above procedure for multitaper spectrogram, we present an example from human EEG under propofol-induced general anesthesia. Fig. 7(a) shows the EEG data from a frontal electrode of a subject undergoing propofol-induced

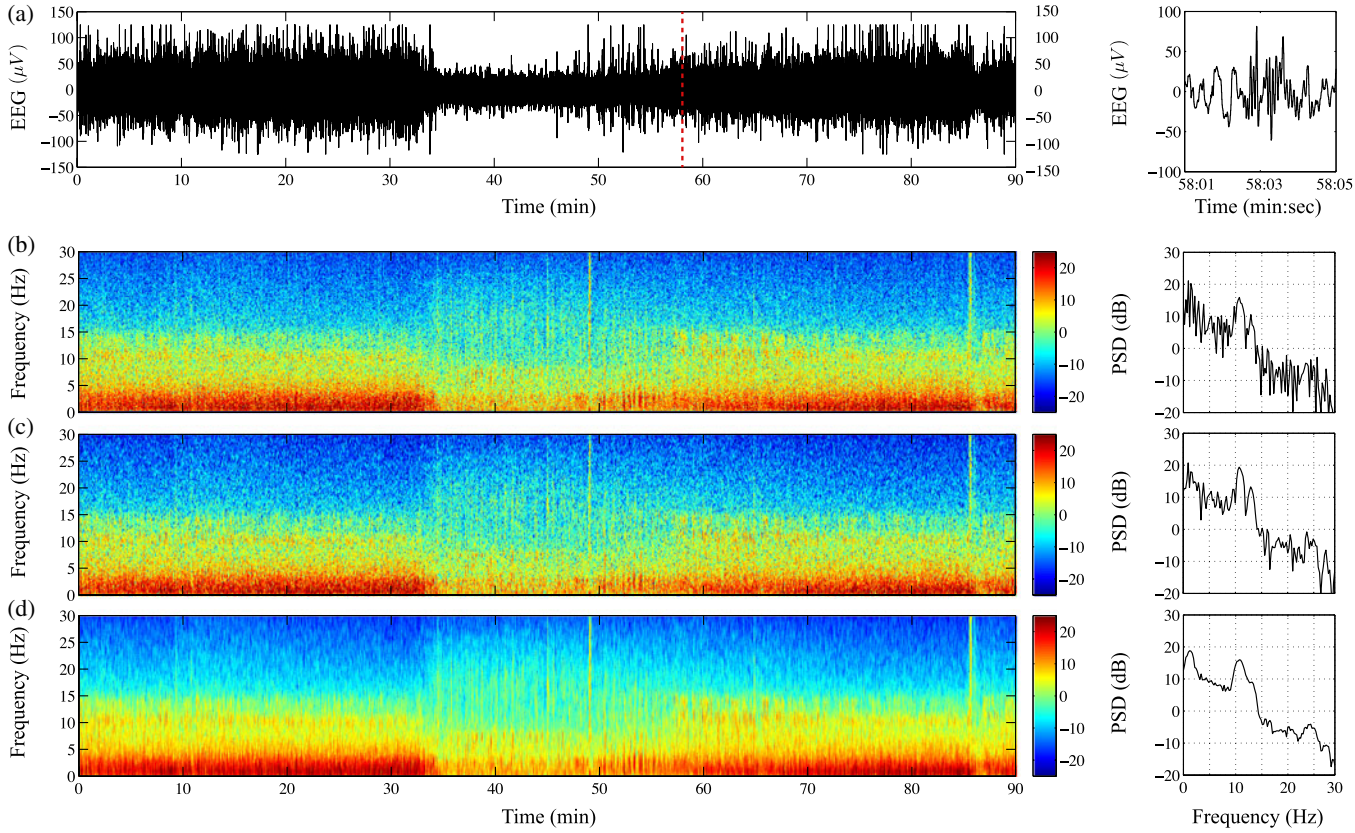

Fig. 8. Spectrogram of EEG during sleep. (a) Time-domain EEG signal, (b) spectrogram from periodogram spectral estimates, (c) spectrogram from tapered spectral estimates using the Hann taper, and (d) spectrogram from multitaper spectral estimates. The right panels show the corresponding zoomed-in view of 4 s of data starting at 58:01 (marked with the red-dashed line in A).

general anesthesia during a surgical procedure lasting ~ 33 min. The anesthetic is administered at around minute 5. The right panel shows the zoomed-in view of 8 s of the data starting at minute 20. The sampling frequency of the EEG is $F_s = \frac{1}{\Delta} = 250$ Hz. The spectrogram representation of the EEG data was computed through the following steps.

1) We estimate the time scale on which the data are stationary to be of the order of a few seconds. We choose $T = 8$s (which implies $N := \lfloor \frac{T}{\Delta} \rfloor = 2000$).
2) We pick a spectral resolution of 1 Hz (which implies $\alpha = \frac{N\Delta R}{2} = 4$).
3) We pick a sliding length of $\tau = 4$s (which implies $K = 1000$).
4) We find the multitaper estimate of the PSD of the data at a window of length 2000 (using four tapers), centered at time indices $n = 1000, 2000, 3000, \ldots$.

Fig. 7(b) shows the spectrogram of the EEG data obtained by the periodogram estimate at each window. The colormap of the PSD is in the dB scale. During propofol-induced general anesthesia, dominant alpha (8–12 Hz) activity arises in the frontal area of the scalp, whereas in the awake state, the frontal alpha activity is highly suppressed [11]. The right panel shows the zoomed in view of the PSD estimate at minute 20, which shows high activity in the alpha band. Fig. 7(c) shows the spectrogram where the Hann taper is used to obtain the spectral estimates. Finally, Fig. 7(d) shows the spectrogram obtained by the multitaper method. Clearly, the spectrograms obtained by periodogram and tapering have higher variances than those obtained by the multitaper method (see the right panels).

As another example, we consider the spectral representation of human EEG data during sleep. The data were obtained from the SHHS Polysomnography Database (available at: http://www.physionet.org/pn3/shhpsgdb/). Fig. 8(a) shows the EEG data from a dorsal electrode of a subject during night sleep. The duration of the data is ~ 1.5 h. The sampling frequency is $F_s = \frac{1}{\Delta} = 250$ Hz. Since during sleep transient events occur on a time scale of a few seconds, we choose $T = 4$s. We also choose a spectral resolution of 2 Hz, since the frequency bands of interest during sleep are typically wider than 2 Hz. Finally, we choose a sliding length of $\tau = 1$ s. The spectrogram representation of the EEG data was computed through the following steps.

1) We estimate the time scale on which the data are stationary to be of the order of a few seconds. We choose $T = 4$ s (which implies $N := \lfloor \frac{T}{\Delta} \rfloor = 1000$).
2) We pick a spectral resolution of 2 Hz (which implies $\alpha = \frac{N\Delta R}{2} = 4$).
3) We pick a sliding length of $\tau = 1$ s (which implies $K = 250$).

4) We find the multitaper estimate of the PSD of the data at a window of length 500 (using four tapers), centered at time indices $n = 500, 750, 1000, \ldots$.

Fig. 8(b) shows the periodogram estimate of the EEG data shown in Fig. 8(a). The colormap of the PSD is in the dB scale. During sleep, delta activity (< 4Hz) as well as transient sleep spindles (~ 1 s waveforms with 8–14-Hz frequency content) [26], [27] are the dominant spectral components, and are indicators of the sleep stage [27], [28]. Up to minute ~ 34, the subject is in the nonrapid eye movement (NREM) sleep stages, where dominant delta activity and sleep spindles are present. From minute ~ 34 to 57, the subject is in the rapid eye movement stage, where the delta activity is significantly reduced and the sleep spindles are absent. From minute ~ 57 on the subject transitions back into the NREM sleep. The right panel zoomed-in view of Fig. 8(a) shows a sleep spindle activity in the middle of the observation window. The right panel of Fig. 8(b) shows the zoomed in view of the PSD estimate which shows high activity in the spindle frequency band. Fig. 7(c) corresponds to the spectrogram obtained using the Hann taper at each window. Finally, Fig. 7(d) shows the spectrogram obtained by the multitaper method at each window. Similar to the case of anesthesia EEG data, the spectrograms obtained by periodogram and tapering have higher variances than those obtained by the multitaper method.

VI. Discussion and Conclusion

Spectral analysis is one of the most commonly used signal processing techniques. Nonparametric spectral analysis based on Fourier methods is appealing because this approach makes it possible to quickly decipher the harmonic structure of complex nonstationary signals using limited *a priori* knowledge. Proper use of these methods does require care and an understanding of potential problems and their solutions. Although multitaper spectral methods were devised more than 30 years ago, they are still not widely used in signal processing analyses in bioengineering. Because multitaper spectral methods provide a principled way to solve the challenging problem of balancing the bias-variance tradeoff in spectral estimation, we felt that these techniques would be a timely topic for this review.

The bias-variance tradeoff is defined by the inequality in (23). Finding the optimal solution for the tradeoff requires minimizing the quadratic form in (24) which is equivalent to finding the eigenvectors of the matrix $\mathbf{\Phi}_R$ in (26). The first L eigenvectors of $\mathbf{\Phi}_R$ defined by its L largest eigenvalues, are the dpss and provide the tapers to be applied to the data.

Application of the multitaper spectral methods is a two-step process. First, as always, the user has to choose the time window over which local stationarity of the signal is assumed to hold. Second, the user must specify the desired level of spectral resolution within the time window. Once the level of spectral resolution is set, then given the sampling interval of the data, the number of tapers which minimizes the bias-variance tradeoff is computed using (27).

The multitaper spectral estimates minimize variance by averaging L approximately independent spectral estimates. The spectral estimates are independent because the tapers are orthogonal. The larger L is chosen then, the smaller the variance of the spectral estimate. However, as L increases bias increases because the degree of side-lobe suppression of the tapers decreases leading to leakage of power from other frequencies (see Fig. 5). Specifying the spectral resolution effectively chooses the width of the central lobe of the equivalent taper. Hence, frequencies that differ by less than the spectral resolution are indistinguishable.

In addition, the statistical properties of multitaper methods are well characterized making it easy to construct confidence intervals about any specific frequency band and also to use hypothesis testing methods to evaluate in a principled way the differences between spectra computed from two different signals (see, for example, [6], [7], and [29]). Although we have not included an analysis of the spectral coherence, it must be noted that cross-spectral quantities can be readily computed using multitaper methods [24]. Multitaper spectral methods are implemented in MATLAB and R. They are also part of the Chronux tool box maintained at website www.chronux.org.

We hope that this review helps enhance the use of multitaper spectral methods in bioengineering signal processing problems.

References

[1] G. Buzsaki, *Rhythms of the Brain*. Oxford, U.K.: Oxford Univ. Press, 2009.

[2] T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2008.

[3] W. J. Emery and R. E. Thomson, *Data Analysis Methods in Physical Oceanography*. New York, NY, USA: Elsevier, 2001.

[4] M. Ghil, M. Allen, M. Dettinger, K. Ide, D. Kondrashov, M. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou, "Advanced spectral methods for climatic time series," *Rev. Geophys.*, vol. 40, no. 1, pp. 3-1–3-41, 2002.

[5] O. Yilmaz, *Seismic Data Analysis: Processing, Inversion, and Interpolation of Seismic Data*. Tulsa, OK: Society of Exploration Geophysicists, 2001.

[6] D. Percival and A. Walden, *Spectral Analysis for Physical Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

[7] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, Sep. 1982.

[8] P. P. Mitra and B. Pesaran, "Analysis of dynamic brain imaging data," *Biophys. J.*, vol. 76, no. 2, pp. 691–708, 1999.

[9] S. T. Williams, M. M. Conte, A. M. Goldfine, Q. Noirhomme, O. Gosseries, M. Thonnard, B. Beattie, J. Hersh, D. I. Katz, J. D. Victor, S. Laureys, and N. D. Schiff, "Common resting brain dynamics indicate a possible mechanism underlying zolpidem response in severe brain injury," *eLife*, vol. 2, 2013.

[10] S. Derégnaucourt, P. P. Mitra, O. Fehér, C. Pytte, and O. Tchernichovski, "How sleep affects the developmental learning of bird song," *Nature*, vol. 433, no. 7027, pp. 710–716, 2005.

[11] P. L. Purdon, E. T. Pierce, E. A. Mukamel, M. J. Prerau, J. L. Walsh, K. F. K. Wong, A. F. Salazar-Gomez, P. G. Harrell, A. L. Sampson, A. Cimenser, S. Ching, N. J. Kopell, C. Tavares-Stoeckel, K. Habeeb, R. Merhar, and E. N. Brown, "Electroencephalogram signatures of loss and recovery of consciousness from propofol," *Proc. Nat. Acad. Sci. USA*, vol. 110, pp. E1142–E1151, 2013.

[12] Q. Noirhomme, R. I. Kitney, and B. Macq, "Single-trial EEG source reconstruction for brain–computer interface," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 5, pp. 1592–1601, May 2008.

[13] H. S. Bokil, B. Pesaran, R. A. Andersen, and P. P. Mitra, "A method for detection and classification of events in neural activity," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 8, pp. 1678–1687, Aug. 2006.

[14] A. M. Yaglom, *Correlation Theory of Stationary and Related Random Functions: Volume 1: Basic Results*. New York, NY, USA: Springer-Verlag, 1987.

[15] C. Chatfield, *The Analysis of Time Series: An Introduction*. Boca Raton, FL, USA: CRC Press, 2003.
[16] A. Schuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena," *Terr. Magn.*, vol. 3, no. 1, pp. 13–41, 1898.
[17] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra: From the Point of View of Communications Engineering*. New York, NY, USA: Dover, 1959.
[18] T. P. Bronez, "Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 36, no. 12, pp. 1862–1873, Dec. 1988.
[19] M. S. Bartlett, "Periodogram analysis and continuous spectra," *Biometrika*, vol. 37, no. 1, pp. 1–16, 1950.
[20] P. D. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967.
[21] K. S. Lii and M. Rosenblatt, "Prolate spheroidal spectral estimates," *Stat. Probab. Lett.*, vol. 78, no. 11, pp. 1339–1348, 2008.
[22] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty—I," *Bell Syst. Tech. J*, vol. 40, no. 1, pp. 43–63, 1961.
[23] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
[24] A. Walden, "A unified view of multitaper multivariate spectral estimation," *Biometrika*, vol. 87, no. 4, pp. 767–788, 2000.
[25] T. P. Bronez, "On the performance advantage of multitaper spectral analysis," *IEEE Trans. Signal Process.*, vol. 40, no. 12, pp. 2941–2946, Dec. 1992.
[26] L. De Gennaro and M. Ferrara, "Sleep spindles: An overview," *Sleep Med. Rev.*, vol. 7, no. 5, pp. 423–440, 2003.
[27] R. W. McCarley, "Neurobiology of REM and NREM sleep," *Sleep Med.*, vol. 8, no. 4, pp. 302–330, 2007.
[28] M. Steriade, D. A. McCormick, and T. J. Sejnowski, "Thalamocortical oscillations in the sleeping and aroused brain," *Science*, vol. 262, no. 5134, pp. 679–685, 1993.
[29] D. Denison, A. Walden, A. Balogh, and R. Forsyth, "Multitaper testing of spectral lines and the detection of the solar rotation frequency and its harmonics," *J. Royal Statist. Soc.: Ser. C (Appl. Statist.)*, vol. 48, no. 4, pp. 427–439, 1999.

Behtash Babadi (M'08) received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2006, and the M.Sc. and Ph.D. degrees in engineering sciences from Harvard University, Cambridge, MA, USA, in 2008 and 2011, respectively.

From 2011 to 2013, he has been a post-doctoral fellow at the Department of Brain and Cognitive Sciences at Massachusetts Institute of Technology, Cambridge, MA as well as at the Department of Anesthesia, Critical Care and Pain Medicine at Massachusetts General Hospital, Boston, MA. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Maryland, College Park. His research interests include statistical and adaptive signal processing, neural signal processing, compressed sensing, and systems neuroscience.

Emery N. Brown (M'01–SM'06–F'08) received the B.A. degree from Harvard College, Cambridge, MA, USA, the M.D. degree from Harvard Medical School, Boston, MA, and the A.M. and Ph.D. degrees in statistics from Harvard University, Cambridge, MA.

He is the Edward Hood Taplin Professor of Medical Engineering in the Institute for Medical Engineering and Science and a Professor of computational neuroscience in the Department of Brain and Cognitive Sciences at Massachusetts Institute of Technology; the Warren M. Zapol Professor of Anaesthesia at Harvard Medical School; and an Anesthesiologist in the Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital. His statistical research focuses on development of signal processing algorithms to study of neural systems. His experimental research uses systems neuroscience approaches to study how mechanisms of general anesthesia.

Dr. Brown is a Fellow of the American Statistical Association, a Fellow of the American Association for the Advancement of Science, a Fellow of the American Academy of Arts and Sciences and a member of the Institute of Medicine. He was a recipient of a 2007 NIH Directors Pioneer Award, the 2011 National Institute of Statistical Sciences Sacks Award for Outstanding Cross-Disciplinary Research, and a 2012 NIH Director's Transformative Research Award. He is a member of the NIH BRAIN Initiative Working Group.