33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Statistical Characterization of Complex-Valued EEG Spectrum During Mental Imagery Tasks

Amirhossein S. Aghaei, Mohammad Shahin Mahanta, Konstantinos N. Plataniotis, and Subbarayan Pasupathy

Abstract— **Electroencephalogram (EEG) recordings of brain activities can be processed in order to augment the brain's cognitive, sensory, or motor functionality. A representative, yet analytically tractable, model is essential to EEG processing. Several studies have examined different statistical models for EEG power spectrum. But recent studies have shown that not only the power, but also the phase of the spectrum, carries relevant information on brain activities. As a result, this paper focuses on the complex-valued spectrum of EEG, and proposes a general non-circularly-symmetric multivariate Gaussian model for this spectrum. This simple model can encapsulate the information in both power and phase of the spectrum, and its validity for EEG data has been verified using standard statistical tests.**

I. Introduction

During the past two decades, electroencephalogram (EEG) signals have been widely used in brain-computer interface (BCI) systems to provide a non-muscular channel for the brain to communicate with the external world. A BCI translates the electrical activity of the brain into signals that control external devices, which can be utilized to help disabled individuals or used in commercial applications. One type of these systems is *spontaneous* BCI, where the EEG signal generated from a mental imagery task, such as hand/foot movement imagination, is used for brain-computer interfacing. This paper focuses on the analysis of EEG signals generated by these mental imagery tasks.

Many feature extraction methods which are currently used for spontaneous BCI systems are based on frequency domain, also called spectral, analysis of EEG signals. Although complete representation of the EEG signal in the frequency domain results in a complex-valued representation, these methods usually only consider the *power spectral density* and ignore the *phase* of EEG spectrum.[1] However, recent studies in neuroscience have revealed that there exist relevant information carried in the phase of electrical activities of the brain, both in microscopic level (the phase of neural firings) and in macroscopic level (the phase of EEG signals) [1]–[4].Furthermore, recent studies on EEG source separation algorithms using independent component analysis (ICA) method have shown that utilization of the complex-valued EEG spectrum, instead of power spectrum, significantly improves the performance of ICA algorithm [5].

The authors are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Canada. {aghaei, mahanta, kostas, pas}@comm.utoronto.ca

[1]Note that the phase of EEG spectrum is different from the phase coupling of oscillatory activities from different parts of the brain, which is usually measured by *phase locking value*.

In this paper, we study the statistical characteristics of the *complex-valued* representation of the EEG signals in the frequency domain. Since EEG characteristics change over time, this complex-valued representation is obtained using a short-time Fourier transformation (STFT), as explained in the next section. In the literature, there exist several studies on the statistical properties of power spectral density of the EEG during different brain activities; however, to the best of our knowledge, there is no statistical study on the complex-valued spectrum of these signals.

We propose a complex-valued multivariate Gaussian model for the EEG spectrum and perform various statistical tests to study how accurately this model fits experimental EEG data. Furthermore, in order to study if there exist any relevant information in the phase of EEG spectrum, we examine its *propriety* or *circular-symmetry* properties. By definition, a complex-valued random variable z is called proper or circularly-symmetric if its real and imaginary parts are independent and have equal power. If z is proper, its phase is uniformly distributed and conveys no information. In many situations, the complex-valued spectrum of the signal is proper, and hence the phase of its spectrum is ignored and only the power of the spectrum is analyzed. In the case of EEG signals obtained during mental imagery tasks, however, our studies reveal that the EEG spectrum is *improper*. As a result, we conclude that the phase of EEG spectrum conveys information relevant to brain activities, which is ignored in the commonly used real-valued power spectral representation. The rest of this paper is organized as follows. Section II provides the general complex-valued multivariate Gaussian model which is used in this paper for the EEG spectrum. Section III explains the specifications of the experimental setup, and Section IV discusses the results for each statistical test used to verify the proposed model.

II. Multivariate Complex Gaussian Model

In this paper, we propose a multivariate Gaussian model for the complex-valued EEG spectrum. We show that like many other biological signals, EEG spectrum can be modeled by a multivariate Gaussian distribution. The main advantage of a Gaussian model is that complete characterization of this model only requires estimation of the first and second order statistics of the data. Furthermore, a multivariate Gaussian model provides a mathematically tractable framework for development of more efficient signal processing and feature extraction algorithms for analysis of the EEG spectrum. In Section IV, we will examine the validity of this model for complex-valued spectrum of the EEG signal in each channel.

A key feature of the model presented in this paper is that here we deal with complex-valued entities which require special second-order treatment, as explained below. Let $\mathbf{z} = [z_1, z_2, \ldots, z_K]^T$ be a Gaussian random vector[2] each element of which is a complex-valued random variable that can be decomposed into its real and imaginary parts as follows: $z_i = x_i + jy_i$. Second order characterization of $\mathbf{z}$ requires knowledge of both of the following matrices [6]:

$$\text{covariance}: \mathbf{C}_{\mathbf{z}\mathbf{z}^H} = \mathrm{E}\left\{(\mathbf{z}-\bar{\mathbf{z}})(\mathbf{z}-\bar{\mathbf{z}})^H\right\} \quad (1)$$

$$\text{pseudo-covariance}: \mathbf{C}_{\mathbf{z}\mathbf{z}^T} = \mathrm{E}\left\{(\mathbf{z}-\bar{\mathbf{z}})(\mathbf{z}-\bar{\mathbf{z}})^T\right\} \quad (2)$$

Let $\tilde{\mathbf{z}} = [x_1, \ldots, x_K, y_1, \ldots, y_K]^T$. Then, $\mathbf{C}_{\mathbf{z}\mathbf{z}^T}$ and $\mathbf{C}_{\mathbf{z}\mathbf{z}^H}$ are uniquely determined by $\mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T}$, and vice versa. Indeed, it can be shown that

$$\mathbf{C}_{\mathbf{z}\mathbf{z}^H} = \mathbf{C}_{\mathbf{x}\mathbf{x}^T} + \mathbf{C}_{\mathbf{y}\mathbf{y}^T} + j(\mathbf{C}_{\mathbf{y}\mathbf{x}^T} - \mathbf{C}_{\mathbf{x}\mathbf{y}^T}), \quad (3)$$

$$\mathbf{C}_{\mathbf{z}\mathbf{z}^T} = (\mathbf{C}_{\mathbf{x}\mathbf{x}^T} - \mathbf{C}_{\mathbf{y}\mathbf{y}^T}) + j(\mathbf{C}_{\mathbf{x}\mathbf{y}^T} + \mathbf{C}_{\mathbf{y}\mathbf{x}^T}), \quad (4)$$

where $\mathbf{x} = [x_1, \ldots, x_K]^T$ and $\mathbf{y} = [y_1, \ldots, y_K]^T$. The probability density function (pdf) of $\mathbf{z}$ can be determined in terms of $\mathbf{C}_{\mathbf{z}\mathbf{z}^H}$ and $\mathbf{C}_{\mathbf{z}\mathbf{z}^T}$ or alternatively in terms of $\mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T}$. Throughout this paper, we use the following formulation for the pdf of vector $\mathbf{z}$ in terms of the vector $\tilde{\mathbf{z}}$:

$$f_{\mathbf{z}}(\mathbf{z}) = |2\pi\mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\tilde{\mathbf{z}}-\bar{\tilde{\mathbf{z}}})^T \mathbf{C}^{-1}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T}(\tilde{\mathbf{z}}-\bar{\tilde{\mathbf{z}}})\right\}.$$

By definition, random vector $\mathbf{z}$ is called *proper* [6] or *circularly symmetric* [7] if $\mathbf{C}_{\mathbf{z}\mathbf{z}^T} = \mathbf{0}$; otherwise, it is called *improper* or *non circularly-symmetric*. From (4), it can be seen that a proper vector $\mathbf{z}$ has the following properties: $\mathbf{C}_{\mathbf{x}\mathbf{x}^T} = \mathbf{C}_{\mathbf{y}\mathbf{y}^T}$ and $\mathbf{C}_{\mathbf{x}\mathbf{y}^T} = -\mathbf{C}^T_{\mathbf{x}\mathbf{y}^T}$. In the univariate case, these conditions will be reduced to the ones explained in Section I. As mentioned there, the phase of a proper complex variable is distributed uniformly and conveys no information. Thus, we will use the propriety of the EEG spectrum to measure whether or not its phase conveys any information.

III. Experimental Setup

In this paper, our analysis is based on data set V of the BCI competition III [8] which is available online. This data set consists of EEG signals of three normal subjects recorded during four non-feedback sessions. During each session, the subject sequentially imagines three different tasks which constitute the different classes: repetitive self-paced *left hand* movements (class 1), repetitive self-paced *right hand* movements (class 2), and *generation of words* beginning with the same random letter (class 3). Each task lasts 15 seconds and is continuously followed by another randomly selected task requested by the operator. The EEG signals are recorded at 512Hz sampling rate using a Biosemi system with 32 electrodes located according to the International 10-20 system. Our analysis is performed on 8 centro-parietal channels: C3, Cz, C4, CP1, CP2, P3, Pz, and P4, which are suggested by the data set providers.

[2]In this paper, scalars are shown in lowercase (e.g., a), column vectors in boldface lowercase (e.g., $\mathbf{a}$), and matrices in boldface uppercase (e.g., $\mathbf{A}$). The transpose and the Hermitian transpose of $\mathbf{A}$ are respectively denoted by $\mathbf{A}^T$ and $\mathbf{A}^H$. Also, $\bar{a} = \mathrm{E}\{a\}$, and $\mathrm{E}\{.\}$ denotes expectation.

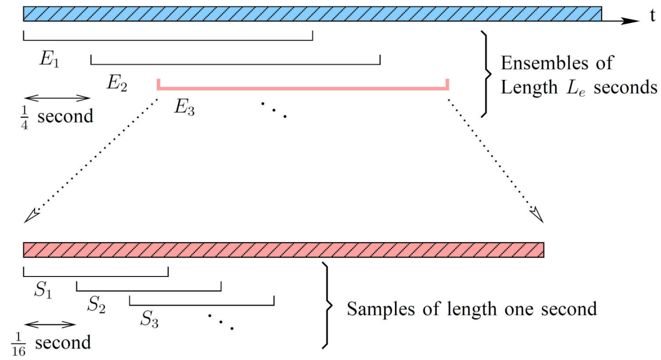


Fig. 1. Observation windows of length L_e and their corresponding samples.

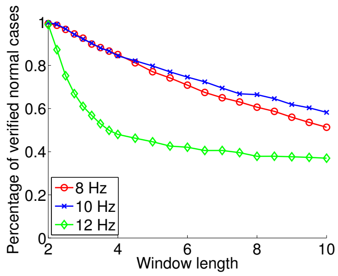
Consider the multichannel EEG signal recorded using these 8 electrodes during a mental imagery task. This signal can be represented by a matrix with the columns corresponding to different channels and the rows corresponding to different time instances. We apply an STFT on each channel of this EEG data to get the frequency domain representation of the EEG at each time instant. The applied STFT uses overlapping Tukey windows of length 1 second ($N_w = 512$) and $\alpha = 1/8$. Then, the spectral components in the range of $8-32$Hz with a frequency resolution of 2Hz are retained. This frequency band approximately corresponds to the α rhythm ($8-12$ Hz) and β rhythm ($12-30$ Hz) of the brain which are known to be associated with mental imagery tasks. The resulting multichannel EEG spectrum at any time can be represented by a complex-valued matrix $\mathbf{Z}$ of size 13×8.

It is well known in the literature that the characteristics of EEG signals during mental imagery tasks are time-varying. However, over a short observation period, EEG can be considered to be quasi-stationary. Assuming that the EEG data is quasi-stationary over an observation window of L_e seconds, we consider all complex-valued EEG spectrums that are observed during this period to form a set of samples with the same statistical characteristics. For simplicity, we call this set of samples an *ensemble*.

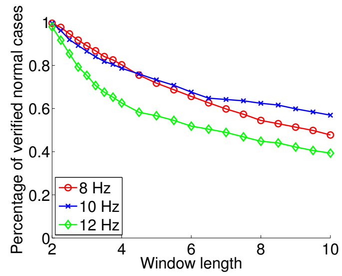
As illustrated in Figure 1, the EEG signal during each mental imagery task is divided into several overlapping observation periods (i.e., $E_1, E_2, \ldots$) of length L_e. During each E_i, the signal is transformed from time-domain to the frequency-domain, using STFT of length one second with overlapping factor of $15/16$. The resulting samples (i.e., $S_1, S_2, \ldots$) form an ensemble E_i. Each multichannel sample (S_i) in this ensemble can be represented by a complex-valued matrix $\mathbf{Z} \in \mathbb{C}^{13\times 8}$, where each column of $\mathbf{Z}$ represents the vector of 13 frequency components of an EEG channel.

We propose that if the duration of the observation window (L_e) is short enough, each column of $\mathbf{Z}$ can be modeled as an improper complex-valued Gaussian random vector as defined in Section II. This model will be verified in three steps:

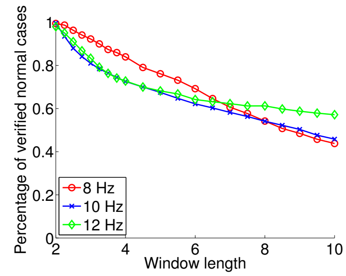
1) Validating the normality of individual components of $\mathbf{Z}$, denoted by z_{mn}, for different values of L_e and finding the maximum value of L_e over which all z_{mn} fit the complex-normal model;
2) Validating the joint-normality of each column vector



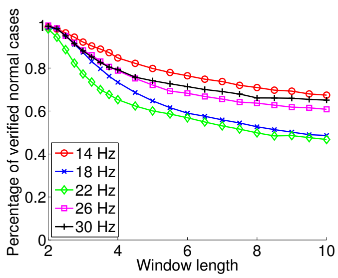
(a) α-band, subject 1



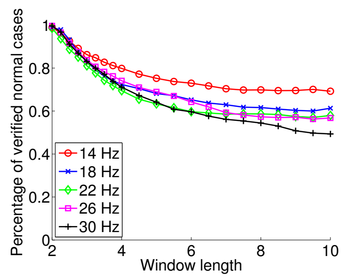
(b) α-band, subject 2



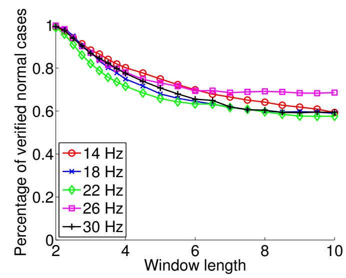
(c) α-band, subject 3



(d) β-band, subject 1



(e) β-band, subject 2



(f) β-band, subject 3

Fig. 2. Percentage of verified normal EEG components for left hand movement task performed by different subjects is plotted for different frequencies averaged over channels. (For more clarity, only five frequency components of β-band are illustrated.)

of $\mathbf{Z}$, denoted by $\mathbf{z}_n$, over the observation length L_e determined in Step 1.

3) Validating the impropriety of each $\mathbf{z}_n$, over the observation length L_e determined in Step 1.

IV. Statistical Tests

A. Testing the Normality of z_{mn}

This section studies the normality of each complex-valued frequency component of the multichannel EEG spectrum $\mathbf{Z}$. We test the following null hypothesis

$$H_0 : z_{mn} = x_{mn} + jy_{mn} \sim \mathcal{CN}(\mu_z, \mathbf{C}_{zz^*}, \mathbf{C}_{zz}),$$

i.e., z_{mn} has a univariate complex-valued Gaussian distribution with unknown mean, variance, and pseudo-variance. As described in Section II, H_0 is equivalent to the hypothesis

$$H_0' : \tilde{\mathbf{z}}_{mn} = \begin{bmatrix} x_{mn} \\ y_{mn} \end{bmatrix} \sim \mathcal{N}_2(\mu_{\tilde{\mathbf{z}}}, \mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T}),$$

i.e., $\tilde{\mathbf{z}}_{mn}$ has a bivariate real-valued Gaussian distribution with unknown mean and covariance.

We examine hypothesis H_0' using the well-known Mardia's multivariate normality test [9] with a significance level of 0.05. In order to find the maximum length L_e over which the EEG signal can be assumed to be quasi-stationary, we have repeated Mardia's test for various values of ensemble length L_e from 2 to 10 seconds. The test results for Task 1 of all three subjects are shown in Figure 2. We have reported the percentage of ensembles whose samples are verified to have Gaussian distribution. Parts (a-c) of this figure illustrate the results for α-band frequency components, and Parts (d-f) illustrate the results for β-band. It should be mentioned that since all the channels exhibited similar trends in the tests, the results reported in Figure 2 have been averaged over all 8 channels. This figure reveals that despite of the inter-subject and inter-frequency variability of the results, in all the situations, the complex-valued Gaussian model describes the experimental data more accurately as the length of L_e decreases. Specifically, for $L_e = 3$ seconds, on average only %15 of the of the ensembles are rejected to have samples with normal distribution. The test results show similar trend for the other two tasks.

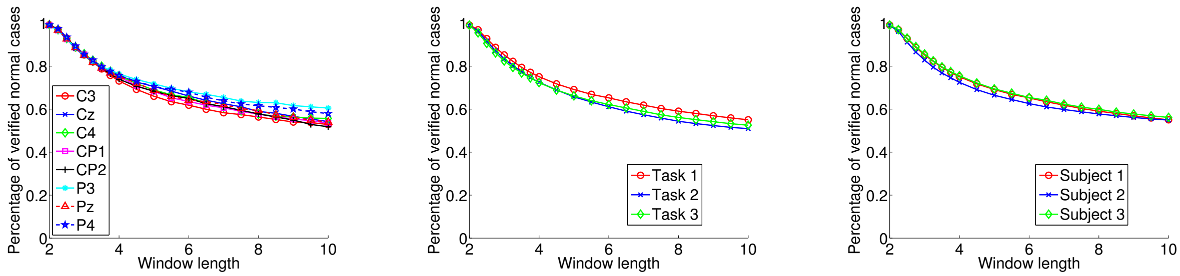
It is worthy to mention that when the resulting percentages are averaged over all frequencies, there is no significant variation between different tasks, different subjects, or different channels. As an example, Figure 3.a provides the average percentage of verified normal cases for first subject's left hand movement task, plotted for all the 8 channels. Figure 3.b compares the results of all three tasks for the first subject, which are again very close to each other. The same trend can be seen in Figure 3.c for one task over all three subjects.

As a result, we can conclude that H_0 is valid if the length of the observation window is small enough. Thus, we set $L_e = 3$ seconds in the rest of this paper. It should be noted that even for large L_e, only in half of the cases H_0 is rejected.

B. Testing the Normality of $\mathbf{z}_n$

The results of the previous test showed each individual z_{mn} element can be modeled as a complex-valued Gaussian random variable when L_e is small enough. This is a necessary but not sufficient condition for *joint* Gaussianity of all elements of vector $\mathbf{z}_n$. This section examines the following hypothesis $H_0 : \mathbf{z}_n = \mathbf{x}_n + j\mathbf{y}_n \sim \mathcal{CN}_M(\mu_{\mathbf{z}}, \mathbf{C}_{\mathbf{z}\mathbf{z}^H}, \mathbf{C}_{\mathbf{z}\mathbf{z}^T})$, where $M = 13$ is the number of frequency components in the EEG spectrum, and $\mathcal{CN}_M$ denotes the M-variate complex-valued Gaussian distribution. H_0 is equivalent to the following hypothesis:

$$H_0' : \tilde{\mathbf{z}}_n = \begin{bmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{bmatrix} \sim \mathcal{N}_{2M}(\mu_{\tilde{\mathbf{z}}}, \mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T}).$$



(a) Left hand movement, First subject (b) First subject, All tasks (c) Left hand movement, All subjects

Fig. 3. Percentage of verified normal EEG components for (a) left hand movement of Subject 1 in different channels, (b) different tasks of the first subject averaged over channels, and (c) left hand movement of different subjects averaged over channels. The values in all figures are averaged over frequencies.

TABLE I

PERCENTAGE OF VERIFIED MULTI-VARIATE COMPLEX NORMAL EEG CHANNELS FOR DIFFERENT TASKS IN DIFFERENT SUBJECTS.

Task	Subj. 1	Subj. 2	Subj. 3
(1)	0.9978	0.9676	0.9899
(2)	0.9488	0.9939	0.9943
(3)	0.9303	0.9873	0.9938

TABLE II

AVERAGE P-VALUE OF MULTI-VARIATE NORMALITY TEST ON EEG CHANNELS FOR DIFFERENT TASKS IN DIFFERENT SUBJECTS.

Task	P_1			P_2		
	Subj. 1	Subj. 2	Subj. 3	Subj. 1	Subj. 2	Subj. 3
(1)	0.9842	0.9740	0.9798	0.3118	0.3019	0.3139
(2)	0.9622	0.9776	0.9811	0.3042	0.3180	0.3135
(3)	0.9500	0.9803	0.9822	0.3010	0.3148	0.3118

Assuming $L_e = 3$ seconds from previous section, we have 32 samples in each ensemble to examine the multivariate vector $\tilde{\mathbf{z}}$ of the relatively large dimension $2M = 26$. Consequently, Mardia's multivariate normality test cannot be utilized in this section. We use the multivariate normality test proposed by [10], which is designed to overcome this small-sample-size problem. The results of this test for a significance value of 0.05 are presented in Table I. The average p-values of this test are also reported in Table II.

The fact that multivariate-normality of $\mathbf{z}_n$ is only rejected in less than $\%10$ of the cases, together with the fact that individual elements of $\mathbf{z}_n$ are shown to be normal, confirms with high confidence that our proposed multivariate normal model for the vector $\mathbf{z}_n$ fits the experimental data.

C. Testing the Propriety of $\mathbf{z}_n$

This section examines the propriety, or circular-symmetry, of the complex-valued EEG spectrum for each channel. As mentioned in Section II, if $\mathbf{z}_n$ is proved to be improper, we can conclude that the phase of complex-valued spectrum obtained from STFT has relevant information which will be lost in the power spectral density representation. Therefore, we examine the hypothesis $H_0 : \mathbf{C}_{\mathbf{x}\mathbf{x}^T} = \mathbf{C}_{\mathbf{y}\mathbf{y}^T},$ which is a necessary condition for $\mathbf{z}_n$ to be proper. In other words, rejection of H_0 is a sufficient condition for $\mathbf{z}_n$ to be improper.

We use the test in [11], which examines the equality of two covariance matrices with small number of samples. The results for a significance value of 0.05 show that for *all* the cases, hypothesis H_0 is rejected. Indeed, all the resulting test statistics have large values, with an average of 165, whereas the critical value for hypothesis rejection is 1.645.

V. CONCLUSION

Motivated by the recent findings in neuroscience, this paper proposed a *complex-valued* multivariate Gaussian model for the EEG spectrum. This model was verified with several statistical tests on a BCI data set. The test results establish that the underlying component-wise and joint Gaussianity assumptions in this model conform with the inherent data structure. Furthermore, a statistical test for propriety of the complex-valued spectrum demonstrated that EEG spectrum is improper or non-circularly-symmetric. This indicates that not only the amplitude but also the phase of the EEG spectrum conveys relevant information. This, in turn, necessitates the use of *complex-valued spectrum* rather than the *power spectrum* for analysis of EEG data.

The results of this paper can be exploited in the development of new feature extraction or classification algorithms for BCI sytems, which take into account all the information conveyed by the complex-valued representation of the EEG signals in the frequency-domain.

REFERENCES

[1] J. Huxter, N. Burgess, and J. O'Keefe, "Independent rate and temporal coding in hippocampal pyramidal cells," *Nature*, vol. 425, no. 6960, pp. 828–832, 2003.

[2] P. Fries, D. Nikolic, and W. Singer, "The gamma cycle," *Trends in neurosciences*, vol. 30, no. 7, pp. 309–316, 2007.

[3] N. A. Busch, J. Dubois, and R. VanRullen, "The phase of ongoing EEG oscillations predicts visual perception," *Journal of Neuroscience*, vol. 29, no. 24, pp. 7869–7876, 2009.

[4] T. Masquelier, E. Hugues, G. Deco, and S. J. Thorpe, "Oscillations, phase-of-firing coding, and spike timing-dependent plasticity: An efficient learning scheme," *Journal of Neuroscience*, vol. 29, no. 43, pp. 13 484–13 493, 2009.

[5] A. Hyvrinen, P. Ramkumar, L. Parkkonen, and R. Hari, "Independent component analysis of short-time fourier transforms for spontaneous EEG/MEG analysis," *NeuroImage*, vol. 49, no. 1, pp. 257–271, 2010.

[6] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, July 1993.

[7] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.

[8] J. Millan, "On the need for on-line learning in brain-computer interfaces," in *Proc. Int. Joint Conf. on Neural Networks*, vol. 4, July 2004, pp. 2877–2882 vol.4.

[9] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.

[10] J. Liang, R. Li, H. Fang, and K.-T. Fang, "Testing multinormality based on low-dimensional projection," *Journal of Statistical Planning and Inference*, vol. 86, no. 1, pp. 129 – 141, 2000.

[11] J. R. Schott, "A test for the equality of covariance matrices when the dimension is large relative to the sample sizes," *Comput. Stat. Data Anal.*, vol. 51, no. 12, pp. 6535–6542, 2007.