Random Matrix Analysis of Human EEG Data

P. Šeba
Department of Physics, University of Hradec Králové, Víta Nejedlého 573, Hradec Králové, Czech Republic and Institute of Physics, Academy of Sciences of the Czech Republic, Cukrovarnická 10, CZ - 162 53 Prague, Czech Republic

(Received 26 February 2003; published 7 November 2003)

We use random matrix theory to demonstrate the existence of generic and subject-independent features of the ensemble of correlation matrices extracted from human EEG data. In particular, the spectral density as well as the level spacings was analyzed and shown to be generic and subject independent. We also investigate number variance distributions. In this case we show that when the measured subject is visually stimulated the number variance displays deviations from the random matrix prediction.

DOI: 10.1103/PhysRevLett.91.198104 PACS numbers: 87.19.Nn, 05.45.Tp

The analysis of the EEG signal has a long history. Being used as a diagnostic tool for nearly 70 years, it still resists a strict and objective analysis and its interpretation remains to be mostly intuitive and heuristic. In particular, the cross-channel relations of the signal and its reference to the sources and hence to certain morphological and/or functional brain structures are not understood sufficiently and are a matter of intense research [1].

Many methods have been developed with the aim of helping to understand the meaning of the EEG signal. They range from a visual inspection of the record by an experienced physician to sophisticated estimations which attempt to describe the signal using phase space methods or to model its sources by a set of oscillating electric dipoles [2]. However none of these methods can be regarded as fully satisfactory.

The reason is that the EEG signal—an electric activity of the brain measured by electrodes (channels) placed on the scalp—is a superposition of electric signals which are produced by a synchronous activity of numerous neurons. The spatial propagation of the electric signal in the complex brain tissue is far from being straightforward. Moreover, various groups of neurons participate in various and sometimes independent tasks. All this together makes the resulting signal complex and difficult to interpret. It has to be stressed that the morphological and functional properties of the brain are not always the same. On the contrary, they can be different for different individuals. This makes the EEG signal dependent on the measured subject. Even if different persons perform the same tasks and the measurement is done under identical conditions the resulting signals could differ.

The aim of this Letter is to demonstrate the existence of features of the EEG signal which are universal, i.e., do not depend on the subject being measured. To this purpose we will investigate statistical properties of the cross-channel correlations of the EEG data and compare them with the predictions of the random matrix theory.

It is known that random matrix theory describes correctly the spectral statistics of certain complex systems. It is useful, in particular, when one deals with systems which are chaotic but display at the same time certain wave (or coherent) properties. A typical example is quantum chaotic systems—see [3] for a review.

An EEG signal is a multivariate time series. Being measured on the scalp, the activity of a particular cerebral area influences the results of several EEG channels. Hence the activity of a given cortex area leads to correlations between the signal measured on different electrodes. The object to be analyzed here is the simplest one describing this correlation, namely, the correlation matrix $C_{l,,m}(T)$ of the signal:

$$C_{l,m}(T) = \sum_{j=N_1}^{N_2} x_l(t_j)x_m(t_j). \quad (1)$$

Here $x_l(t_j)$ denotes the EEG signal measured in the channel l at time t_j and the sum runs over $j = N_1, N_1 + 1, \ldots, N_2$ such that $t_j \in (T, T + \Delta)$. The length Δ of the time interval used to evaluate the correlation matrix will be set to 150 ms here. We assume that during this time interval the activity of the electric sources in the brain can be considered as constant. The obtained signal was not filtered with a single exception, namely, a notched filter serving to remove the 50 Hz noise induced by the electric supply. The signal $x_l(t_j)$ used to evaluate (1) was additionally preprocessed so that its mean is set to zero and the variance equal to 1,

$$\sum_{j=N_1}^{N_2} x_l(t_j) = 0; \qquad \sum_{j=N_1}^{N_2} x_l(t_j)^2 = 1. \quad (2)$$

The importance of correlations for the signal analysis was recognized a few years ago by Kwapien and collaborators when they investigated the MEG (magnetoencefalography) signal of a brain auditory response [4]. In addition to the correlation matrix they used also a more elaborated concept of mutual information to extract the cross-channel correlations.

 0031-9007/03/91(19)/198104(4)$20.00

Our aim is to analyze the EEG correlations using tools of the random matrix theory. To this aim it is necessary to define a matrix ensemble based on the EEG data. Here we use the fact that the brain is nonstationary. This means that the tasks it is performing change with the time. Hence the structure of the correlation matrix (1) changes. A typical EEG measurement lasts about 15–20 min. This time interval can me divided into roughly 7000 not overlapping stationary windows over which the correlation matrix (1) is evaluated. In such a way we get a set of 7000 correlation matrices. These matrices will represent the ensemble we will work with.

Because of the individuality of the EEG signal and also due to the problems related to the technique of attaching the electrodes to the scalp we cannot mix together data obtained during different measurement sessions. This means that even for one person we get different measurements for different matrix ensembles. There is one more drawback which is related to the above definition of the 150 ms wide stationary window: to get enough sampling points inside a particular window we have to use measurements performed with a high sampling rate (we will use data obtained with the sampling rate of 1012 Hz).

Random matrix theory deals with statistical properties of the eigenvalues ξ_n of the correlation matrix C,

$$C|\xi_n\rangle = \xi_n|\xi_n\rangle, \tag{3}$$

with $|\xi_n\rangle$ being the corresponding eigenvectors. The simplest property of the eigenvalues family related to a matrix ensemble is the eigenvalue density $\rho(\xi)$. It counts the mean number of eigenvalues contained in an interval (a, b), i.e.,

$$\{\xi_n; \xi_n \in (a, b)\} = \int_a^b \rho(\xi) d\xi. \tag{4}$$

Assuming that the time series $x_l(t_j)$ is random and Gaussian for all l, i.e., that for each l the values $x_l(t_j)$ are random, normally distributed with the variance equal to 1 and not correlated, we get an ensemble of random matrices which was mathematically studied by Marchenko and Pastur [5,6]. In particular, the density $\rho(\xi)$ of this ensemble is known and given by the formula [7]

$$\rho(\xi) = \frac{\sqrt{(\xi_{\max} - \xi)(\xi - \xi_{\min})}}{2\pi Q\xi}, \tag{5}$$

where $Q = M/N$ with M being the number of channels, N is the number of the sampling points t_j, $N > M$, and $\xi_{\max} = 1 + Q + 2\sqrt{Q}$; $\xi_{\min} = 1 + Q - 2\sqrt{Q}$.

The spectral density is known to be dependent on the underlying data. Since EEG is not a random signal, but on the contrary, it is synchronized and correlated by the corresponding brain activity, we might expect that the spectral density will not follow the Marchenko-Pastur prediction and will be subject dependent. This is, to our surprise, not true at all. We have investigated the EEG records of 90 people and evaluated the corresponding level density function. The result is plotted in Fig. 1 (to keep the plot simple we displayed densities evaluated for three different subjects as a typical example).

For small eigenvalues the spectral density indeed depends on the measured subject. It displays nevertheless a profound and subject-independent algebraic tail for large ξ. This is a surprise since one would expect just the opposite behavior. Specifically, one might expect that small eigenvalues of the correlation matrix feel the influence of the system noise. The spectral density should therefore display universal behavior for small eigenvalues similar to that of the Marchenko-Pastur ensemble. The large eigenvalues on the other hand contain the information about significant correlations and hence about processes in the brain. Consequently, one would expect a nonuniversal and subject-dependent behavior.

Instead of being described by the Marchenko-Pastur ensemble the spectral density obtained with the help of the EEG signal seems to fit better into the predictions valid for Wishard matrices and chiral ensembles [3]. For these ensembles, which are related to the covariance matrices of the signal, the spectral density has an algebraic singularity at the origin, just like the density derived from the EEG data.

The universal behavior of the spectral density for large ξ points to the fact that the seemingly individual brain activity contains some common level of synchronization which is reflected in the generic large tail behavior of the spectral density. This fact remains valid even when the subjects are stimulated (we tried acoustic and optical stimulations). The algebraic character of the tail is also of importance. A similar behavior (universal and algebraiclike tail of the spectral density) has been observed

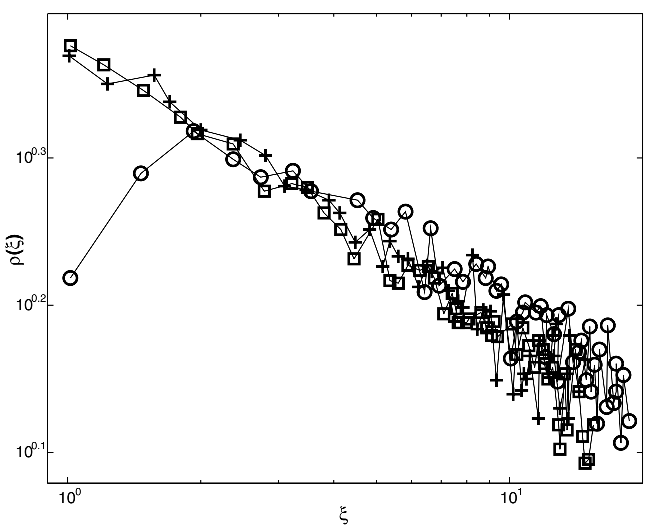

FIG. 1. The eigenvalue density is plotted for data obtained by measuring three different people under the same conditions. The behavior for large ξ is similar in all three cases and shows a clear algebraic behavior (a straight line in the log-log plot).

also for correlation matrices of systems with local fluctuations governed by Lévy flights [8]. A typical example displaying this type of behavior is a time series describing the price fluctuations of individual stocks at an asset market. It has been argued by Burda and collaborators [9] that the existence of an algebraic tail in the spectral density is a sign for critical behavior of the underlying system. Using MEG measurements, signs of critical behavior of the brain have been found also in the power spectra of the related signal which has a "$1/f$" type behavior [10]. This was interpreted as an indication for self-organized critical behavior of the brain which is generated by a fractal (scale-invariant) avalanchelike mechanism.

In the principal-component method the eigenvalues and eigenvectors of the correlation matrix are used to identify the most important sources of the signal. Small eigenvalues are usually neglected and the related correlations are regarded as unimportant. But the fact that some eigenvalue is small does not necessarily mean that the related signals are not important. The "smallness" of the eigenvalue can be a consequence of external influence which screens a part of the signal coming from a certain domain. To investigate the spectral statistics of the correlation matrices more closely we have to first put all the eigenvalues on the same footing. This is done by unfolding the spectra, i.e., by a spectral mapping $\xi_n \to \tilde{\xi}_n$ that "unifies" the system-dependent spectral densities in such a way that the resulting eigenvalue density of $\tilde{\xi}_n$ is constant,

$$\tilde{\xi}_n = N(\xi_n) \tag{6}$$

with $N(E)$ being the spectral counting function

$$N(E) = \int_0^E \rho(\xi) d\xi. \tag{7}$$

After the unfolding we get $\rho(\tilde{\xi}_n) = 1$.

In the remaining part of this Letter we will work with unfolded eigenvalues without mentioning them explicitly and we will omit the tilde in the notation for the sake of simplicity. The random matrix theory predicts universal statistical properties of the unfolded spectra provided the underlying matrix ensemble is large enough to sufficiently fill the space of all matrices with a given symmetry. This generic behavior is observed, for instance, in the spectra of chaotic quantum systems. To demonstrate that a similar universal property is observed also for correlation matrices resulting from the EEG measurements we will focus on two statistical distributions widely used in the field of quantum chaos.

The simplest one, which takes into account only correlations between neighboring eigenvalues, deals with spacings $s_n = \xi_{n+1} - \xi_n$ (here we assume the eigenvalues ξ_n to be ordered with respect to the magnitude, i.e., $\xi_n \leq \xi_{n+1}$). The probability density of finding a given spacing s in the spectra is for random real symmetric matrices well approximated by the Wigner formula

$$P(s) = \frac{\pi}{2} s \exp\left(-\frac{\pi}{4} s^2\right). \tag{8}$$

Another, more subtle distribution that takes into account simultaneous correlation between a group of L subsequent eigenvalues is the number variance Σ^2 defined as

$$\Sigma^2(L) = \langle(\{\xi_n < L\} - L)^2\rangle \tag{9}$$

(here $\{\xi_n < L\}$ denotes the number of eigenvalues smaller then L).

In this case the theory of random matrices predicts

$$\Sigma^2(L) \approx \frac{2}{\pi^2}\left(\log(2\pi L) + 1.5772 - \frac{\pi^2}{8}\right). \tag{10}$$

These distributions are of universal validity. This has been confirmed, for instance, for spectra of chaotic wave systems that display this distribution independently of their physical character. A similar observation has been done also for a time series which characterizes the evolution of atmospheric quantities [11] or the price fluctuations on an asset marker [12].

As we have already mentioned the EEG signal is subject dependent. Nevertheless, the spectral statistical properties of the correlation matrices are universal. In particular, the behavior of the level spacing distribution and number variance obtained from the EEG data of healthy persons are well described by Eqs. (8) and (10), respectively.

We have investigated the data of 90 people which were measured without an external stimulation. A part of the data was obtained by a standard clinical EEG device with

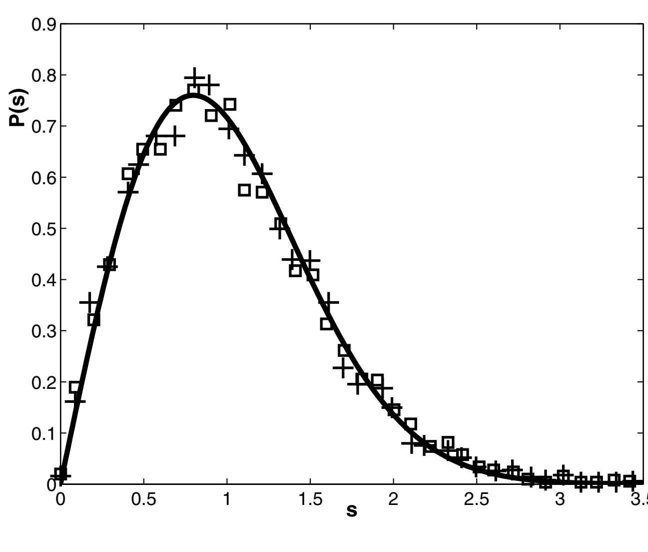

FIG. 2. A typical level spacing distribution obtained for a resting subject (eyes open, no stimulation) (crosses) and a stimulated subject (squares). The result is compared with formula (8) (solid line).

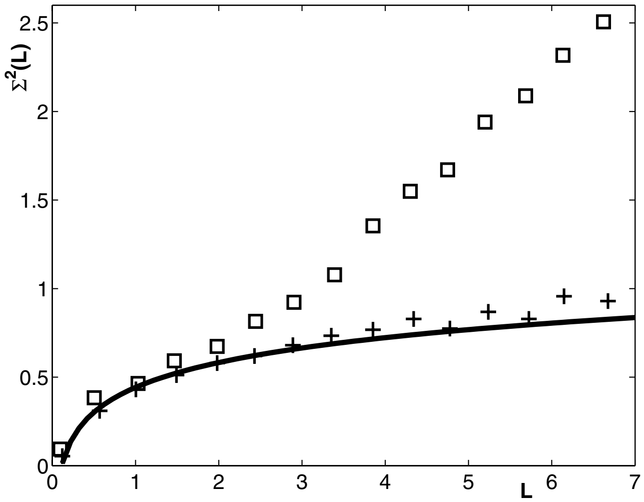

FIG. 3. A typical number variance obtained for a resting subject (eyes open, no stimulation) (crosses) and a stimulated subject (squares). The result is compared with formula (10) (solid line).

19 channels and a part with the help of an experimental device having 15 channels which was placed inside a Faraday trap to avoid external influence. Before the analysis the data were visually inspected and parts corrupted with artifacts (eyes blinking, a motion of the subject, etc.) were removed. Finally, the resulting spectral data were unfolded. As a result we obtained a convincing universal behavior which is in accordance with the predictions of the random matrix theory.

Typical results for the level spacing and number variance distributions are shown in Figs. 2 and 3. The data were obtained under the circumstances that volunteers were measured first under resting conditions (no stimulation) and then again when these persons were visually stimulated by a periodically moving square. The stimulation is expected to change the correlation pattern of the data due to its large response in the visual cortex. Nevertheless, it turned out that these changes are too subtle to influence the level spacing distribution. In all cases we find a very good agreement with the random matrix theory prediction. The standard deviation between the values given by the Wigner formula and the results obtained from the EEG data is equal to 0.026 for the resting subject and 0.027 for the stimulated case.

The number variance is, however, quite sensitive and changes when the subject is visually stimulated. The random matrix prediction is in this case valid for nonstimulated data only, where the standard deviation between the predicted and measured values is equal to 0.05. (The measurement with a stimulation leads to the standard deviation as large as 0.6.) This can be easily understood. A visual stimulation changes the activity of the visual cortex which proceeds the visual input. As a consequence, its relative correlations with the remaining parts of the cortex are lowered which leads finally to the observed deviation of $\Sigma^2(L)$ for larger L.

This result is of interest since the number variance may change not only due to an external stimulation but also when the correlation ensemble is influenced, for instance, by some pathological process. It has to be stressed at this point, however, that number variance is also quite sensitive to artifacts. To obtain a result which is in agreement with the random matrix theory, all artifacts have to be carefully removed. We mean by artifacts those changes in the EEG signal that lead within the correlation window to a variance which is 2 times larger than the mean variance evaluated over all correlation windows. Such windows were rejected and were not used in further considerations.

This work was supported by the Grant No. 202/02/0088 of the Czech Grant Agency. A cooperation with the Institute of Pathological Physiology in Hradec Králové and the neurological department in Rychnov nad Kneznou is also gratefully acknowledged.

[1] F. Cincotti *et al.*, Methods Inf. Med. **41**, 337 (2002).
[2] M. Scherg, in *Fundamentals of Dipole Source Potential Analysis*, edited by F. Grandori, M. Hoke, and G. L. Romani, Advances in Audiology Vol. 6 (S. Karger, Basel, Switzerland, 1990), p. 40.
[3] T. Guhr, A. Mueller-Groeling, and H. A. Weidenmueller, Phys. Rep. **299**, 189 (1998).
[4] J. Kwapien, S. Drozdz, and A. A. Ioannides, Phys. Rev. E **62**, 5557 (2000).
[5] V. A. Marchenko and L. A. Pastur, Mat. Sb. **72**, 507 (1967).
[6] L. A. Pastur, Russian Math. Surveys **28**, 1 (1973).
[7] F. Gotze and A.Tikhomirov, Probab. Theory Appl. **47**, 381 (2002).
[8] Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed, cond-mat 0103140.
[9] Z. Burda, J. Jurkiewicz, M. A. Nowak, G. Papp, and I. Zahed, cond-mat 0103109.
[10] S. Drozdz, J. Kwapien, A. A. Ioannides, and L. C. Liu, cond-mat/9901134.
[11] M. S. Santhanam and P. K. Patra, Phys. Rev. E **64**, 16 102 (2001).
[12] B. Rosenov, V. Plerou, P. Gopikrishnan, A. Luis, N. Amaral, and H. E. Stanley, Int. J. Theor. Appl. Finance **3**, 399 (2000).