Data Mining in EEG Wave Trains in Early Stages of Parkinson's Disease

Olga S. Sushkova[1(✉)], Alexei A. Morozov[1,2], and Alexandra V. Gabova[3]

[1] Kotel'nikov Institute of Radio Engineering and Electronics of RAS, Mokhovaya 11-7, Moscow 125009, Russia
o.sushkova@mail.ru, morozov@cplire.ru

[2] Moscow State University of Psychology and Education, Sretenka 29, Moscow 107045, Russia

[3] Institute of Higher Nervous Activity and Neurophysiology of RAS, Butlerova 5A, Moscow 117485, Russia

Abstract. A method of analysis and visualization of electroencephalograms (EEG) based on wave trains is developed. In this paper, we use the "wave train" term to denote a signal localized in time, frequency, and space. The wave train is a typical pattern in a background EEG and detecting/analyzing such signals gives useful information about the brain activity. Alpha spindles, beta spindles, and sleep spindles are the best known examples of the wave trains in EEG. The preliminary results of the research give evidence that EEG analysis and visualization method based on the wave trains is useful for looking for group statistical regularities in early stages of Parkinson's disease and searching EEG features that are prospective for early stages of Parkinson's disease diagnostics. The novelty of this work is in developing a new method to search parameters of wave trains useful for evaluating the group statistical regularities in EEG data.

1 Introduction

The method of analysis of electroencephalograms (EEG) based on wave trains that was described in [1] was extended and visualization of the data was added. In physics, a wave train (or a wave packet) is a short "burst" or "envelope" of localized wave action that travels as a unit. We use the "wave train" term to denote a signal localized in time, frequency, and space. The wave train is a typical pattern in a background EEG and detecting/analyzing such signals gives useful information about the brain electrical activity. Alpha spindles, beta spindles, and sleep spindles are the best known examples of the wave trains in EEG. Several methods based on Fourier spectra, wavelets, autoregressive models, adaptive filtering, etc. have been developed for detecting and analyzing these EEG patterns (see surveys in [2,3]).

The idea of the new method of EEG analysis consists of that the wave trains in a wide frequency band from 2 Hz to 30 Hz are detected and analyzed using ROC curves. The wave trains are detected as local maxima in a wavelet spectrogram of EEG record. Various attributes of these wave trains are computed.

O. Pichardo-Lagunas and S. Miranda-Jiménez (Eds.): MICAI 2016, Part II, LNAI 10062, pp. 403–412, 2017.
DOI: 10.1007/978-3-319-62428-0_32

Then visualization and statistical analysis on the base of these attributes are implemented. In the paper [1], it was demonstrated that the investigation of wave train attributes reveals the new statistically significant regularities in a group of de novo Parkinson's disease patients. Namely, the Mann-Whitney statistical test indicates a significant decrease of the quantity of the wave trains in the C3 and C4 electrodes in the beta frequency range. In this paper, the problem of searching such wave train features that are appropriate for recognition/diagnosis of early stages of Parkinson's disease on the base of EEG data is considered. Thus, the main difference between this paper and the previous work [1] is in that earlier we have analyzed the wave train features in standard theta, alpha, beta, and gamma frequency ranges and in this paper we also determine frequency ranges that are prospective for the recognition/diagnosis using ROC curves.

The idea of the wave trains in EEG is described in Sect. 2. In the Sect. 3, the experimental setting used for the verification of the wave train analysis method is considered. The data mining method is introduced in Sect. 4. Examples of data visualization are considered in Sect. 5.

2 The Idea of the Wave Trains in EEG

Let M be a local maximum in a wavelet spectrogram (see Fig. 1). We estimate the full width at half maximum (FWHM) of M in the time plane $FWHM_{TIME}$ and in the frequency plane $FWHM_{FREQUENCY}$. Then we check whether there are no values in the rectangle area

$$FWHM_{TIME} \times FWHM_{FREQUENCY}$$

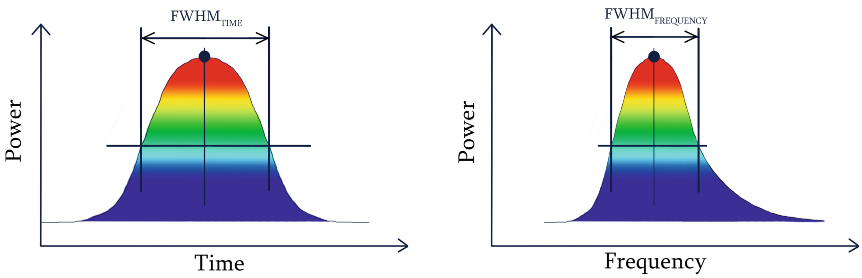

Fig. 1. An example of a spectrogram of a wave train in a time-frequency domain. The diagram on the left shows the spectrogram of the signal in the time plane. The abscissa indicates a time and the ordinate indicates a power. The diagram on the right shows the spectrogram of the signal in the frequency plane. The abscissa indicates a frequency and the ordinate indicates a power.

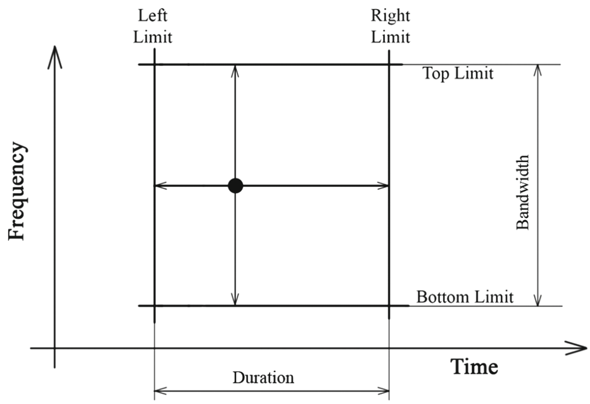

Fig. 2. Time and frequency bounds of the M wave train in the wavelet spectrogram. The abscissa indicates the time and the ordinate indicates the frequency.

that are bigger than the M value (see Fig. 2). We consider M as a case of a wave train if $FWHM_{TIME}$ of M is greater or equal to the T_D threshold (see Fig. 1). The T_D threshold is a function of the central frequency f of the maximum M:

$$FWHM_{TIME} \geq T_D = N_P/f,$$

where N_P is a constant given by an expert. In this paper, we apply the value: $N_P = 2$.

3 The Experimental Setting

In the paper, the analysis and visualization method is considered by an example of real data acquired in a neurophysiological examination. Let us consider a set of wave trains detected in a group of de novo Parkinson's disease patients and a healthy volunteers group (see Fig. 3). The group of patients includes 17 patients with right-hand tremor and 11 patients with left-hand tremor in the first stage of Parkinson's disease without Parkinson's disease treatment. The group of healthy volunteers includes 15 people.

The ages of patients were from 38 to 71 years old; the mean age was 60 years old. The ages of healthy volunteers were from 48 to 81 years old; the mean age was 58 years old. No statistically significant differences between the patients' ages and the volunteers' ages were detected. The amount of male patients was 11; the amount of female ones was 17. The amount of male healthy volunteers was 5; the amount of female ones was 10. The size of the groups is typical for a neurophysiological examination. It is difficult to collect more patients without any treatment in the early stage of Parkinson's disease.

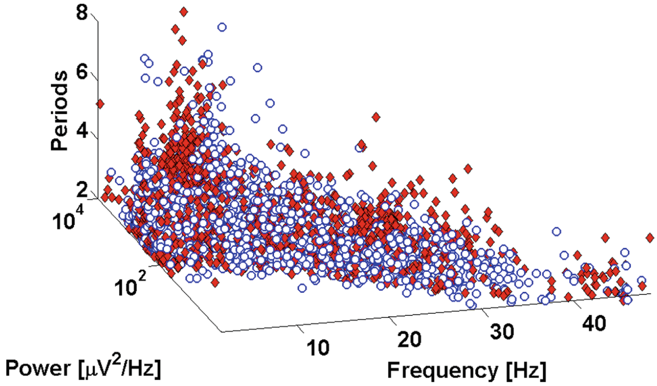

Fig. 3. The set of wave trains detected in the group of de novo Parkinson's disease and the healthy volunteers group. The abscissa indicates frequencies of the wave trains. The ordinate indicates squared amplitudes of the wave trains in a logarithmic scale. The applicate indicates durations of the wave trains. The patients are indicated by dark red diamonds and the healthy volunteers are indicated by light circles. The patients with right-hand tremor, background EEG, channel C4. (Color figure online)

The patients were diagnosed according to the standard Hoehn and Yahr scale. All patients and volunteers were right-handed. A standard 10×20 EEG acquisition schema was used. A background EEG was recorded in standard condition. Examined person sat in an armchair relaxing with arms disposing on the armrests and fingers dangling freely from the ends of armrests. Besides, EEG was recorded in a non-standard condition [4]. Namely, a subject was instructed to keep a special pose with the muscle tension that provokes a tremor: the arms were placed on armrests of the chair; the palms were straightened, placed in a vertical plane, and stretched a bit; the feet were stretched a bit and touched the floor by the heels only. The eyes were closed during the recordings. A 41-channel digital EEG system Neuron-Spectrum-5 (Neurosoft Ltd.) was used for the data acquisition. The sampling rate was 500 Hz. The 0.5 Hz high-pass filter and the 50 Hz supply-line filter were used. Three EEG records were acquired for every subject with interruptions for a rest and a relaxation. The duration of every record was about 2 min. Then the best of three records was selected that contains a minimal number of artifacts. The record was analyzed as is, without selection of areas in the signal.

Special software was developed for analyzing the data. The analysis includes the following EEG pre-processing operations:

1. The Huber's X84 method [5] for outlier rejection was used for removing EEG artifacts.
2. A set of notch filters was applied for removing a power-line noise at 50, 100, 150, and 200 Hz.

3. The eight order 2–240 Hz band pass Butterworth filter was applied. Signals were filtered in the forward and reverse directions to eliminate a phase distortion.
4. Signals were decimated with the decimation factor 4.

The spectrograms were created using the complex Morlet wavelet:

$$\Psi(x) = \frac{1}{\sqrt{\pi F_b}} exp\left(2i\pi F_c x\right) exp\left(-\frac{x^2}{F_b}\right)$$

In this paper, F_b equals 1 and F_c equals 1. The frequency step in the spectrograms equals 0.1 Hz.

In this paper, the C3 and C4 electrodes are considered only, because these electrodes approximately correspond to the motor cortex areas and are situated in the scalp area that produces a minimal number of muscle artifacts.

4 The Data Mining Method

Let $MinFreq$, $MaxFreq$ are frequency bounds of 3D area S (a parallelepiped S) in the space of the wave trains. Let $MinPower$, $MaxPower$ are power bounds of the area S; $MinDurat$, $MaxDurat$ are duration bounds of the area S; and $MinBandwidth$, $MaxBandwidth$ are bandwidth bounds of the area S. One can calculate a number of wave trains per second located in the area S in every individual patient and healthy volunteer and create histograms of the quantity of the wave trains per second (see an example in Fig. 4). A statistical difference between the diagrams may indicate that the area S contains wave trains that are typical for Parkinson's disease, but not for the control group, or vice versa. A second interesting issue is whether one can specify a threshold (a limit of the number of the wave trains in the area S) that separates adequately the histograms, because the presence of such threshold means that the quantity of the wave trains in the area S may be used for the clinical diagnosis of Parkinson's disease. For instance, there is a strong statistical difference between the histograms in the Fig. 4 (the Mann-Whitney test, $p < 0.009$). The diagram demonstrates that a typical number of the wave trains in the control group is about 0.13 per second in the given frequency band. At the same time, a typical number of the wave trains in the patients is about 0.06.

Let us consider a threshold Q of the wave trains quantity in the histograms. Let us utilize this threshold as an indicator of Parkinson's disease, that is, a quantity of the wave trains per second greater than Q in a person N in the given area S will indicate that the person N is probably ill. Thus, the True Positive Rate (TPR) indicates the number of the patients that were diagnosed properly using the threshold Q; and the False Positive Rate (FPR) indicates the number of the controls that were mistakenly diagnosed as Parkinson's patients. A ROC

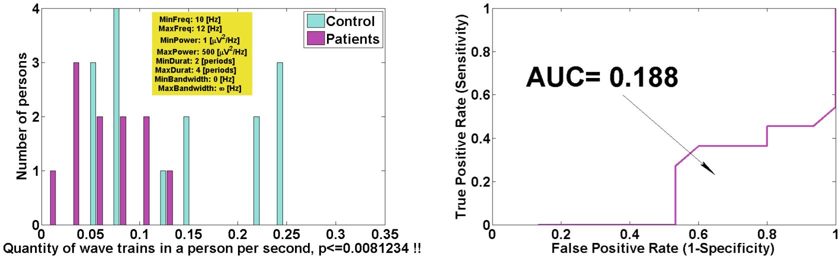

Fig. 4. On the left: histograms of the quantity of the wave trains per second in the patients and the control group. Left hand tremor patients, channel C3. The wave trains are considered in a parallelepiped S bounded by the following limits: a frequency range is 10–12 Hz, a power range is 1–500 μV^2/Hz, the duration range is 2–4 periods. The patient histogram is indicated by the dark magenta color; and the control histogram is indicated by the light cyan color. On the right: a ROC curve based on the histograms. The abscissa indicates the False Positive Rate. The ordinate indicates the True Positive Rate. The area under the ROC curve (AUC) indicates whether the area S is applicable for separation of the patients and the control group. $AUC < 0.5$ indicates that the wave trains quantity is greater in the control group than in the patients. (Color figure online)

curve can be created based on the histograms of the wave trains quantity, that is, a diagram that indicates a relation between the TPR and FPR for various values Q (see Fig. 4). The area under the ROC curve (AUC) is a standard indicator of the quality of the ROC curve; big values of AUC (that are much more than 0.5) and small values of AUC (that are much less than 0.5) indicate that the given area S may be prospective for making clinical diagnosis.

Thus, in mathematical terms, the goal of our investigation is searching such areas in the multidimensional space of the wave trains, where AUC differs sufficiently from 0.5 and is approached to 1 or to 0. $AUC > 0.5$ indicates that the wave trains quantity is greater in the patients than in the controls. Similarly, $AUC < 0.5$ indicates that the wave trains quantity is greater in the control group. An exhaustive search of the values $MinFreq$, $MaxFreq$, $MinPower$, $MaxPower$, $MinDurat$, $MaxDurat$, $MinBandwidth$, and $MaxBandwidth$ can be implemented to investigate the multidimensional space, but we prefer an accurate consideration of different slices of the space using various 2D and 3D diagrams not to miss any interesting regularities in the space of the wave trains. Several examples of this analysis are considered below.

Let us compute AUC values for various frequency ranges. In Fig. 5, the functional dependence of AUC is shown, where the arguments of the function are the $MinFreq$ and $MaxFreq$ bounds. The frequency values varied from 2 to 25 Hz (with the 0.5 Hz step); the $MinPower$, $MaxPower$, $MinDurat$, $MaxDurat$, $MinBandwidth$, and $MaxBandwidth$ were constant: $MinPower =$ 1, $MaxPower = \infty$, $MinDurat = 0$, $MaxDurat = \infty$, $MinBandwidth = 0$,

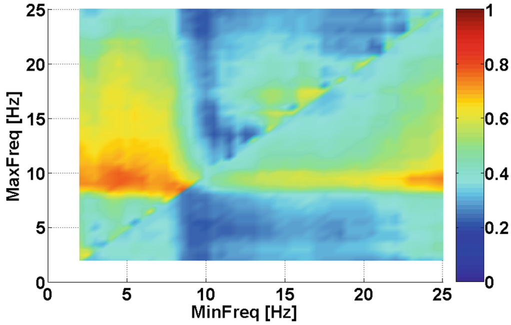

Fig. 5. A diagram of AUC values calculated for various frequency bands. In the upper left triangle of the diagram: the abscissa is the lower bound of the frequency band and the ordinate is the upper bound of the frequency band. In the lower right triangle of the diagram: the abscissa is the upper bound of the excluded frequency band and the ordinate is the lower bound of the excluded band. The frequency varied from 2 to 25 Hz with the 0.5 Hz step. The background EEG was analyzed, right hand tremor patients, the C3 electrode. (Color figure online)

$MaxBandwidth = \infty$. The upper left triangle of the diagram indicates the values of AUC corresponding to the $MinFreq \ldots MaxFreq$ frequency range. The lower right triangle of the diagram indicates the AUC values corresponding to the total frequency band 2 . . . 25 Hz except the $MaxFreq \ldots MinFreq$ band. Note that in the lower right triangle of the diagram the $MinFreq$ indicates the upper limit (but not the lower) of the excepted frequency band and the $MaxFreq$ indicates the lower bound of the excepted values. The diagram in Fig. 5 reveals three interesting frequency ranges that may be prospective for research. The first range is mu (a blue region, the 10.5 . . . 13.5 Hz frequency band approximately), the second is mu too (a red region, the 6 . . . 9.5 Hz frequency band approximately), and the third range is beta (a dark blue region, 18 . . . 24 Hz frequency band approximately). We suppose that the first and second regions in the diagram give a visual evidence for a shift of the mu rhythm to the lower frequency areas. The third region in the diagram is a visual confirmation of the regularity (a significant decrease of the wave train number in the beta band) discovered by a group statistical analysis and reported in [1]. Moreover, a "good" values $AUC = 0.22$ and $AUC = 0.75$ in the first and second colored spots indicate that the mu frequency range may be prospective for making clinical diagnostic.

5 Examples of Data Visualization

Let us consider several examples of AUC diagrams. In Fig. 6, diagrams for left hand tremor patients (the left column) and right hand tremor patients (the right column) are demonstrated. The first row corresponds to the C3 electrode and the

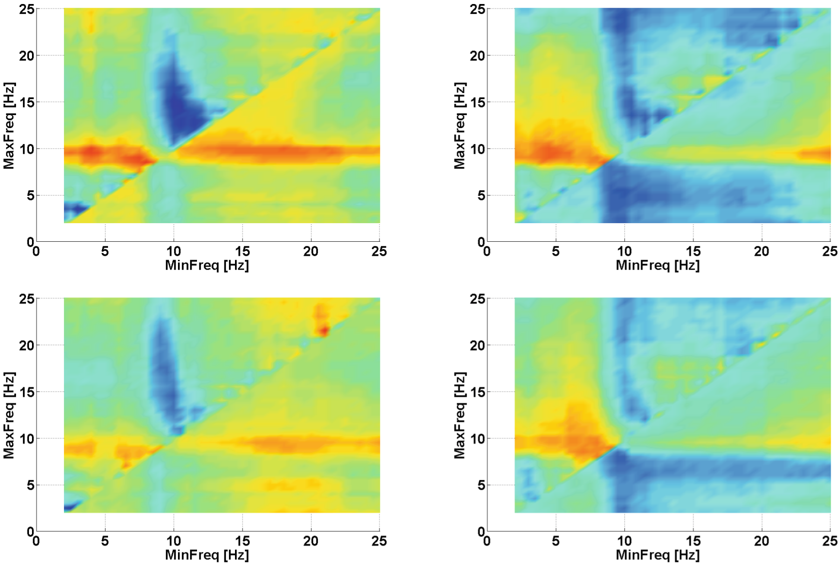

Fig. 6. Diagrams of *AUC* values calculated for various frequency bands. In the left column: left hand tremor patients; in the right column: right hand tremor patients. The first row corresponds to the C3 electrode and the second row corresponds to the C4 electrode. (Color figure online)

second row corresponds to the C4 electrode. One can see a blue area 10 . . . 11 Hz in all the diagrams. The corresponding red areas exist in all the diagrams too. Thus, similar regularities are presented in both C3 and C4 electrodes, but the diagrams corresponding to the left tremor patients and the right tremor patients differ more than the diagrams corresponding to the C3 and C4 electrodes. This is evidence than the compensatory mechanisms in the cortex in these two groups of patients are not equal.

Let us compare these *AUC* diagrams corresponding to the background EEG with diagrams corresponding to the EEG records created in the special pose that provokes tremor. In the Fig. 7, the left column corresponds to the background EEG and the right column corresponds to the special pose EEG. Only the right tremor patient diagrams are demonstrated in both columns. The first row corresponds to the C3 electrode and the second row corresponds to the C4 electrode. Note that in the special pose the mu rhythm colored spots are less pronounced. This is the evidence that we observe a shift of the mu rhythm, but not of the alpha rhythm in the diagrams, because the special pose may suppress the mu rhythm. Note that the blue spot in the beta frequency range is more pronounced in the special pose diagrams. Probably the special pose increases the regularity [1] in the beta frequency range.

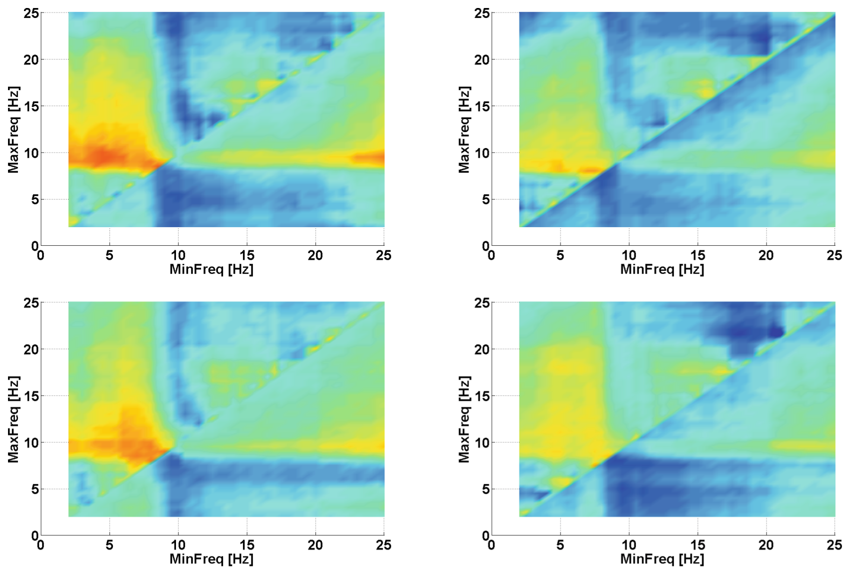

Fig. 7. Diagrams of AUC values calculated for various frequency bands. In the left column: background EEG records; in the right column: the special pose records. Only the right hand tremor patients are analyzed. The first row corresponds to the C3 electrode and the second row corresponds to the C4 electrode.

6 Conclusions

The method of a brain electrical activity investigation based on the EEG wave train analysis and visualization is developed. The preliminary results of the research give evidence that EEG analysis method based on the wave trains is prospective for:

- Looking for group statistical regularities in the early stages of Parkinson's disease that gives a basic knowledge about the disease and compensatory mechanisms in the cortex;
- Searching EEG features that are prospective for the early stages of Parkinson's disease diagnostics.

Acknowledgments. Authors are grateful to Alexei V. Karabanov for selection and medical examination of the patients, to Galina D. Kuznetsova and Alexander F. Polupanov for co-operation and a help in the research, and Yuriy V. Obukhov for a help in the statement of the problem. We acknowledge a partial financial support from the Russian Foundation for Basic Research, grants 16-37-00426 and 15-07-07846.

References

1. Sushkova, O., Morozov, A., Gabova, A.: A method of analysis of EEG wave trains in early stages of Parkinson's disease. In: International Conference on Bioinformatics and Systems Biology (BSB-2016), Allahabad, India, Indian Institute of Information Technology (2016)
2. Parekh, A., Selesnick, I., Rapoport, D., Ayappa, I.: Sleep spindle detection using time-frequency sparsity. In: IEEE Signal Processing in Medicine and Biology Symposium, Philadelphia, PA, pp. 1–6. IEEE (2014)
3. O'Reilly, C., Nielsen, T.: Automatic sleep spindle detection: benchmarking with fine temporal resolution using open science tools. Front. Hum. Neurosci. **9**(353) (2015)
4. Andreeva, Y., Khutorskaya, O.: EMGs spectral analysis method for the objective diagnosis of different clinical forms of Parkinson's disease. J. Electromyogr. Clin. Neurophysiol. **36**, 187–192 (1996)
5. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: Robust Statistics: The Approach Based on Influence Functions. Wiley, New York (1986)