# Finding Significant Correlates of Conscious Activity in Rhythmic EEG

**Piotr J. Durka**

*Laboratory of Medical Physics, Institute of Experimental Physics, Warsaw University, ul. Hoża 69, 00-681 Warsaw, Poland*
*Email: durka@fuw.edu.pl*

One of the important issues in designing an EEG-based brain-computer interface is an exact delineation of the rhythms, related to the intended or performed action. Traditionally, related bands were found by trial and error procedures seeking maximum reactivity. Even then, large values of ERD/ERS did not imply the statistical significance of the results. This paper presents complete methodology, allowing for a high-resolution presentation of the whole time-frequency picture of event-related changes in the energy density of signals, revealing the microstructure of rhythms, and determination of the time-frequency regions of energy changes, which are related to the intentions in a statistically significant way.

**Keywords and phrases:** time-frequency, adaptive approximations, matching pursuit, ERD, ERS, multiple comparisons.

## 1. INTRODUCTION

Thinking of a "brain-computer interface" (BCI), one can imagine a device which would directly process all the brains output—like in a perfect virtual reality machine [1]. Today's attempts are much more humble: we are basically at the level of controlling simple left/right motions. On the other hand, these approaches are more ambitious than direct connections to the peripheral nerves: we are trying to guess the intention of an action directly from the activity of the brains cortex, recorded from the scalp (EEG).

Contemporary EEG-based BCI systems are based upon various phenomena like, for example, visual or P300 evoked potentials, slow cortical potentials, or sensorimotor cortex rhythms [2]. The most attractive path leads towards the detection of the "natural" EEG features, for example a normal intention of moving the right hand (or rather its reflection in EEG) would move the cursor to the right. Determination of such features in EEG is more difficult than using evoked or especially trained responses. Desynchronization of the $\mu$ rhythm is an example of a feature correlated not only with the actual movement, but also with its mere imagination.

All these approaches encounter obstacles, common in the neurosciences: great intersubject variability and poor understanding of the underlying processes. Significant improvement can be brought by coherent basic research on the EEG representation of conscious actions. This paper presents two methodological aspects of such research.

(i) High-resolution parameterization and feature extraction from the EEG time series. Scalp electrodes gather signal from many neural populations, so the rhythms of interest are buried in a strong background. Owing to the high temporal resolution of EEG and the oscillatory character of most of its features, we can look for the relevant activities in the time-frequency plane.

(ii) Determination of *significant* correlates of conscious activities requires a dedicated statistical framework. Until recently, reporting significance of changes in the time-frequency plane presented a serious problem.

## 2. TIME-FREQUENCY ENERGY DENSITY OF SIGNALS

Among the parameters used in nowadays BCI systems (like those designed in the Graz University of Technology [3]), event-related desynchronization and synchronization (ERD/ERS) phenomena play an important role. ERD and ERS are defined as the percentage of change of the average (across repetitions) power of a given rhythm—usually $\mu/\alpha$, $\beta$, and $\gamma$ [4]. Estimation of the time course of the rhythm energy is crucial for the sensitivity of these parameters. But due to the intersubject variability, we cannot expect the rhythms to appear at the same frequencies for all subjects.

Therefore, a classical procedure was developed to find the reactive rhythms [4]. For each subject, the frequency range of interest was divided into 1 Hz intervals, in each of them the single trials (repetitions) were bandpass filtered, squared, and averaged, to obtain the estimate of the average band energy. Among these fixed bands, those revealing the largest changes related to the event were chosen. This naturally limits the
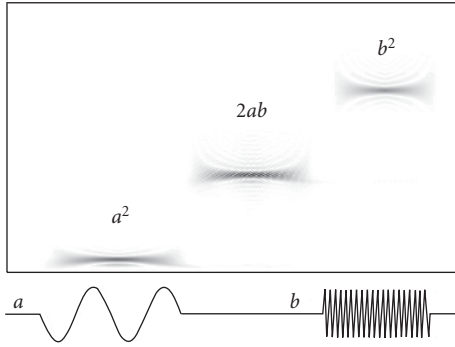
FIGURE 1: Top: Wigner distribution ((A.5); vertical—frequency, horizontal—time) of the signal simulated as two short sines (bottom). We observe the autoterms $a^2$ and $b^2$ corresponding to the time and frequency spans of the sines, and cross-term $2ab$ at time coordinates where no activity occurs in the signal.

frequency resolution to 1 Hz—not taking into account the accuracy of bandpass filtering of finite sequences.

The whole problem is naturally embedded in the time-frequency space. Time-frequency density of signal energy, averaged across trials, provides all the information about the rhythms and the time course of their energy in one clear picture (Figure 2).

### 2.1. Time-frequency distributions of energy density

Because of the uncertainty principle, there are many alternative estimates of the time-frequency density of signal's energy. Actually, the same problem (nonunique estimates) is present also in calculating the spectral power or bandpass filtering finite sequences, but in the quadratic time-frequency distributions we may say that the relevancy of the problem is "squared." Fluctuations of power spectra, appearing at high resolutions, in the time-frequency distributions take the form of cross-terms. These false peaks occur in between the autoterms (which correspond to the actual signal's structures), and significantly blur the energy estimates (Figure 1). Their presence stems from the equation $(a + b)^2 = a^2 + b^2 + 2ab$. Quadratic representation of an unknown signal $s$, composed of two structures $a$ and $b$, contains autoterms corresponding to these structures ($a^2$ and $b^2$) as well as the cross-term $2ab$. For a signal more complex than a sum of two clear and separate structures (like the simplistic simulation in Figure 1), cross-terms are indistinguishable from the autoterms. Advanced mathematical methods are being developed for the reduction of this drawback [5]. While some of them give impressive results for particular signals, in general we are confronted with the tradeoff: higher resolution versus more reliable (suppressed cross-terms) estimate.

### 2.2. Adaptive approximations

If we knew *exactly* the structures ($a$ and $b$) of which the signal is composed, we might explicitly omit the cross-term $2ab$,

thus obtaining a clear time-frequency picture. In practice, this would require a reasonably sparse approximation of the signal in a form

$$s \approx \sum_{n=1}^{M} w_i g_i, \qquad (1)$$

where $g_i$ are known functions fitting well the actual signal's structures. This may be achieved only by choosing the functions $g_i$ for each analyzed signal separately.[1] Criterion of their choice is usually aimed at explaining the maximum part of signal energy in a given number of iterations ($M$). However, the problem of choosing the optimal set of functions $g_i$ is intractable.[2] A suboptimal solution can be found by means of the matching pursuit (MP) algorithm [7]. But even this suboptimal solution is still quite computer-intensive,[3] so the first practical applications were not possible before mid-nineties [8]. The MP algorithm and construction of an estimate of the signal's time-frequency energy density, which is free of cross-terms, are described in the appendix. Functions $g_i$ are chosen from large and redundant collections of Gabor functions (sine-modulated Gauss).

Advantages of this estimator in the context of event-related desynchronization and synchronization were discussed in [9, 10].

## 3. MICROSTRUCTURE OF THE EEG RHYTHMS

### 3.1. Experimental data

To present advantages of the presented methodology, the classical ERD/ERS experimental setup was modified to obtain relatively long epochs of EEG between the events.

Thirty-one-year-old right-handed subject was half lying in a dim room with open eyes. Movement of the thumb, detected by a microswitch, was performed approximately 5 seconds (at a subject's choice) after a quiet sound generated approximately every 20 seconds. Experiment was divided into 15-minute sessions, and recorded EEG into 20-second long. After artifacts rejection, 124 epochs were left for the analysis. EEG was registered from electrodes at positions selected from the 10–20 system. Figures 2–4 present results for the C4 electrode (contralateral to the hand performing movements) in the local average reference. Signal was down-sampled offline from 250 Hz to 125 Hz.

Figure 5 presents data from another subject, collected in a standard ERD/ERS experiment.

---

[1]Contrary to most of the approaches, where all the signals are represented via products with the same set of functions (e.g., basis).

[2]Finding the subset of $M$ functions, which explains the largest ratio of the signal energy among all the other $M$-subsets of the highly redundant set, requires checking *all* the possible $M$-subsets, which leads to the combinatorial explosion even for moderate sets of candidate functions. Problems of such computational complexity are termed NP-hard [6].

[3]Recent results indicate possibilities of a significant decrease of computation times of bias-free MP decompositions.
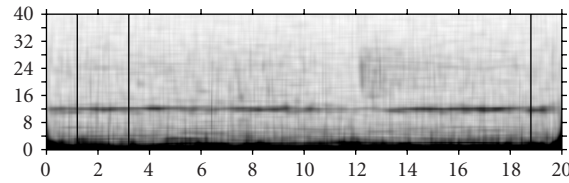
FIGURE 2: Average time-frequency energy density of 124 trials (Section 3.1, energy cut above 2%, sqrt scale); darker area marks higher values of the energy density. Horizontal scale in seconds, vertical in Hz. Finger movement in the 12th second.
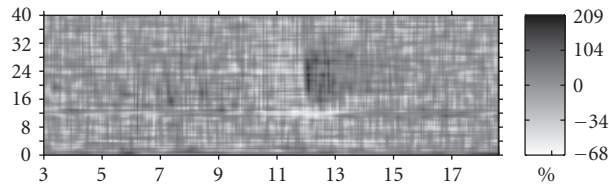


FIGURE 3: ERD/ERS map corresponding to the time between 3 and 19 seconds (vertical lines in Figure 2). Shades of gray are proportional to the percentage of change relative to the reference epoch (between 1 and 3 seconds in Figure 2).

### 3.2. High-resolution picture of energy density

Time-frequency estimates of the signal energy density, including the MP estimate given by (A.5), contain no phase information, so they can be summed across the trials to give the average time-frequency density of energy.[4] Figure 2 presents such an average for 124 repetitions of EEG synchronized to the finger movement, occurring in the 12th second. We easily observe that the $\alpha$ rhythm concentrates around 12 Hz. We may also notice its decrease (desynchronization) around the time when finger movement occurred, as well as some increased activity in 15–30 Hz near 12–13 seconds ($\beta$ synchronization).

In another experiment (Figure 5), high-resolution estimate revealed clearly two very close but separate components of the $\mu$ rhythm with different time courses—an effect elusive to the previously applied methods.

### 3.3. High-resolution ERD and ERS

Speaking of the decrease in the $\alpha$ rhythm in the previous section, we compared the activity near the 12th second (Figure 2) to the average level of the $\alpha$ rhythm energy, or, more correctly, to a period before the movement, which should not be related to the event. To quantify this procedure, we must define the reference period, to which the energy changes will be related. It should be distant enough from the onset of the event, to avoid incorporating premovement correlates into the reference. To avoid border problems of estimates, it should be also removed from the very start of the

---

[4]Note that the average of the energy densities is in general different from the energy density of the averaged signal. The latter (averaged signal) reveals phase-locked phenomena like for example the classical evoked potential.

analyzed epoch. In Figure 2 it was chosen between the 1st and the 3rd second.

Classically, for each selected band, ERD/ERS were calculated as the percentage of power relative to the reference epoch (ERD corresponding to a decrease and ERS to an increase). Owing to the high-resolution estimate of the whole picture of energy density, we may calculate it for the whole relevant time-frequency region with maximum resolution. ERD/ERS map in Figure 3 was obtained as a ratio of each point's energy to the average energy of the reference epoch in the same frequency. In this plot we observe, like in Figure 2, darker area (increase) corresponding to the $\beta$ postmovement synchronization, and white spot around the time of the movement, corresponding to the $\alpha$ desynchronization. However, in the long premovement period there are still a lot of fluctuations, which naturally implies a question about the statistical significance of the observed changes.

## 4. STATISTICAL SIGNIFICANCE

The following steps constitute a fully automatic (hence objective and repeatable) and statistically correct procedure, which delineates and presents with high resolution the time-frequency regions of significant changes in the average energy density.

(1) Divide the time-frequency plane into *resels* (from *res*olution *el*ements), for which the statistics are calculated (Section 4.1).

(2) Calculate pseudo-$t$ statistics and $p$-values for the null hypothesis of no change in the given resel compared to the reference epoch in the same frequency (Section 4).

(3) Select a threshold for the null hypothesis corrected by multiple comparisons (Section 4.3).

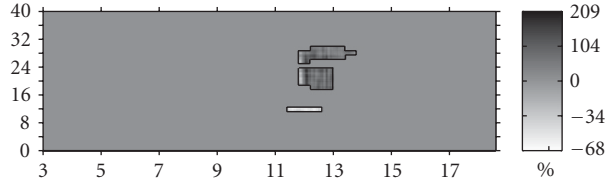(4) Display the energy changes calculated for maximum

FIGURE 4: ERD/ERS from **Figure 3** displayed in regions revealing statistically significant changes in resampling pseudo-$t$ tests (Section 4.2), corrected by 5% false detection rate (Section 4.3).
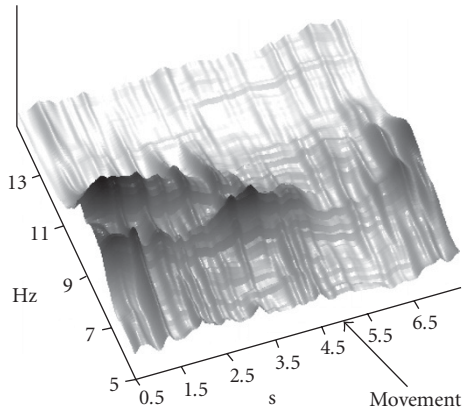


FIGURE 5: Average time-frequency energy density (2) of 57 trials from the C1 electrode (average reference), constructed for $g_{\gamma_i}$ longer than 250 milliseconds. Presented from 5 to 15 Hz is the finger movement in the 5th second. We observe two very close, but separate, $\mu$ rhythms with different time courses. Faster rhythm desynchronizes about 1.5 seconds before the movement, while the slower lasts until its very onset and desynchronizes in the 5th second.

resolution (Section 3.2) in windows corresponding to resels which indicated statistically significant changes.

These steps will be described in the following sections. Further details can be found in [10].

### 4.1. Integration of MP maps in resels

In choosing the dimensions of a resel, suitable for the statistical analyses, we turn to the theory of the periodogram sampling [11]. For a statistically optimal sampling of the periodogram the product of the frequency interval and signal length gives 1/2. This value was taken as the product of the resel's widths in time and frequency, their ratio being a free parameter.

Calculating the amount of energy in such relatively large resels simply as the value of the distribution (A.5) in its center, that is,

$$E_{\text{point}}(t_i, \omega_i) = \sum_n |\langle R^n f, g_{\gamma_n} \rangle|^2 \mathcal{W} g_{\gamma_n}(t_i, \omega_i), \qquad (2)$$

may not be representative for the amount of energy contained in a given resel. In such case[5] we use the exact solution:

$$E_{\text{int}}(t_i, \omega_i)$$
$$= \sum_n |\langle R^n f, g_{\gamma_n} \rangle|^2 \int_{t_i - \Delta t/2}^{t_i + \Delta t/2} \int_{\omega_i - \Delta \omega/2}^{\omega_i + \Delta \omega/2} \mathcal{W} g_{\gamma_n}(t, \omega) dt \, d\omega. \qquad (3)$$

### 4.2. Resampling the pseudo-$t$ statistics

The values of energy of all the $N$ repetitions (trials) in each questioned resel will be compared to the energies of resels within the corresponding frequency of the reference epoch. We denote the time indices $t_i$ of resels belonging to the reference epoch as $\{t_i, i \in \text{ref}\}$ and their number contained in each frequency slice as $N_{\text{ref}}$. For each resel at coordinates $\{t_i, \omega_i\}$ we will compare its energy averaged over $N$ repetitions with the energy averaged over repetitions in resels from the reference epoch in the same frequency. Their difference can be written as

$$\Delta E(t_i, \omega_i) = \frac{1}{N} \sum_{k=1}^{N} E_{\text{int}}^k(t_i, \omega_i) +$$
$$- \frac{1}{N \cdot N_{\text{ref}}} \sum_{k=1}^{N} \sum_{j \in \text{ref}} E_{\text{int}}^k(t_j, \omega_i) \qquad (4)$$
$$= \overline{E(t_i, \omega_i)} - \overline{E(t_{\text{ref}}, \omega_i)},$$

where the superscript "$k$" denotes the $k$th repetition (out of $N$).

However, we want to account also for the different variances of $E^k$, revealing the variability of the $N$ repetitions. Therefore we replace the simple difference of means (4) by the pseudo-$t$ statistics:

$$t = \frac{\Delta E(t_i, \omega_i)}{s_\Delta}, \qquad (5)$$

where $\Delta E$ is defined as in (4), and $s_\Delta$ is the pooled variance of the reference epoch and the investigated resel. In spite of the

---

[5]The difference between (2) and (3) is most significant for structures narrow in time or frequency relative to the dimensions of resels.

central limit theorem, this magnitude tends to have nonnormal distribution [10]. Therefore, we use resampling methods.

We estimate the distribution of $t$ from (5)—under the null hypothesis of no significant change—from the data in the reference epoch (for each frequency $N \cdot N_{ref}$ values) by drawing with replacement two samples of sizes $N$ and $N \cdot N_{ref}$ and calculating, for each such replication, statistics (5). This distribution is approximated once for each frequency. Then for each resel the actual value of (5) is compared to this distribution yielding $p$ for the null hypothesis.

The number of permutations giving values of (5) exceeding the observed value has a binomial distribution for $N_{rep}$ repetitions with probability $\alpha$.[6] Its variance equals $N_{rep}\alpha(1 - \alpha)$. The relative error of $\alpha$ will be then (cf. [12])

$$\frac{\sigma_\alpha}{\alpha} = \sqrt{\frac{(1 - \alpha)}{\alpha N_{rep}}}. \tag{6}$$

To keep this relative error at 10% for a significance level $\alpha = 5\%$, $N_{rep} = 2000$ is enough. Unfortunately, due to the problem of multiple comparisons discussed in Section 4.3, we need to work with much smaller values of $\alpha$. In this study $N_{rep}$ was set to $2 \cdot 10^6$, which resulted in relatively large computation times.

### 4.3. Adjustment for multiplicity

In the preceding section, we estimated the achieved significance levels $p$ for null hypotheses of no change of the average energy in each resel, compared to the reference region in the same frequency. Adjusting results for multiplicity is a very important issue in case of such a large amount of potentially correlated tests. As proposed in [10], it can be effectively achieved using the false discovery rate (FDR, [13]). It controls the ratio $q$ of the number of the true null hypotheses rejected to all the rejected hypotheses. In our case this is the ratio of the number of resels, to which significant changes may be wrongly attributed, to the total number of resels revealing changes.

Let us denote the total number of performed tests, equal to the number of questioned resels, as $m$. If for $m_0$ of them the null hypothesis of no change is true, then [13] proves that the following procedure controls the FDR at the level $q(m_0/m) \leq q$.

(1) Order the achieved significance levels $p_i$, approximated in the previous section for all the resels separately, in an ascending series: $p_1 \leq p_2 \leq \cdots \leq p_m$.

(2) Find

$$k = \max\left\{ i : p_i \leq \frac{i}{m \sum_{j=1}^{m}(1/j)} q \right\}. \tag{7}$$

(3) Reject all hypotheses for which $p \leq p_k$.

---

[6]For brevity we omit the distinction between the exact value $\alpha$, which would be estimated from all the possible repetitions, and the actually calculated.

### 4.4. Display of the statistically significant ERD/ERS

Figure 4 gives the final picture of statistically significant changes in the time-frequency plane. It is constructed by displaying the high-resolution ERD/ERS map (Figure 3) only in the areas corresponding to the resels which revealed statistical significance in the procedure from Section 4. Desynchronization of 12-Hz $\alpha$ occurs around the time of the movement (12th second). Synchronization of 18–30 Hz $\beta$, occurring just after the movement, is divided in half by the harmonic of $\alpha$ (24 Hz). In the long premovement epoch no significant changes are detected, which suggests the robustness and reliability of the whole procedure.

## 5. CONCLUSIONS

Presented procedure gives high-resolution and free-of-crossterms estimates of the average time-frequency energy density of event-related EEG, revealing the microstructure of rhythms. Time-frequency area of significant changes are assessed via objective statistical procedures. This allows for example to investigate the minimum number of repetitions required to delineate the reactive rhythms. Application of this methodology may bring a significant improvement in basic research on the event-related changes of EEG rhythms, as well as "per subject" customization of the ERD/ERS-based BCI.

## 6. REPRODUCIBLE RESEARCH

Software for calculating the MP decomposition (appendix), with complete source code in C and executables for GNU/Linux and MS Windows, plus an interactive display and averaging of the time-frequency maps of energy (in Java), are available at http://brain.fuw.edu.pl/~durka/software/mp. Datasets used in Figures 5–4 and Matlab code for calculating maps and statistics like Figures 2–4: are available at http://brain.fuw.edu.pl/~durka/tfstat/.

## APPENDIX

### MATCHING PURSUIT ALGORITHM

In each of the steps, a waveform $g_{\gamma_n}$ from the redundant dictionary $D$ is matched to the signal $R^n f$, which is the residual left after subtracting results of previous iterations:

$$
\begin{aligned}
R^0 f &= f, \\
R^n f &= \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n} + R^{n+1} f, \\
g_{\gamma_n} &= \arg\max_{g_{\gamma_i} \in D} \left| \langle R^n f, g_{\gamma_i} \rangle \right|,
\end{aligned}
\tag{A.1}
$$

where $\arg\max_{g_{\gamma_i} \in D}$ means the $g_{\gamma_i}$ giving the largest value of the product $|\langle R^n f, g_{\gamma_i} \rangle|$.

Dictionaries ($D$) for time-frequency analysis of real signals are constructed from real Gabor functions:

$$g_\gamma(t) = K(\gamma)e^{-\pi((t-u)/s)^2} \sin\left(2\pi\frac{\omega}{N}(t-u) + \phi\right). \quad \text{(A.2)}$$

$N$ is the size of the signal, $K(\gamma)$ is such that $\|g_\gamma\| = 1$, $\gamma = \{u, \omega, s, \phi\}$ denotes parameters of the dictionary's functions. For these parameters no particular sampling is *a priori* defined. In practical implementations we use subsets of the infinite space of possible dictionary's functions. However, any fixed scheme of subsampling this space introduces a statistical bias in the resulting parameterization. A bias-free solution using stochastic dictionaries, where parameters of the dictionary's functions are randomized before each decomposition, was proposed in [14].

For a complete dictionary the procedure converges to $f$—in theory in an infinite number of steps [7], but in practice we use finite sums:

$$f \approx \sum_{n=0}^{M} \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n}. \quad \text{(A.3)}$$

From this decomposition we can derive an estimate $Ef(t, \omega)$ of the time-frequency energy density of signal $f$, by choosing only autoterms from the Wigner distribution

$$\mathcal{W}f(t, \omega) = \int f\left(t + \frac{\tau}{2}\right)\overline{f\left(t - \frac{\tau}{2}\right)}e^{-i\omega\tau}\, d\tau, \quad \text{(A.4)}$$

calculated for the expansion (A.3). This representation will be a priori free of cross-terms:

$$Ef(t, \omega) = \sum_{n=0}^{M} |\langle R^n f, g_{\gamma_n} \rangle|^2 \mathcal{W}g_{\gamma_n}(t, \omega). \quad \text{(A.5)}$$

## ACKNOWLEDGMENTS

## REFERENCES

[1] S. Lem, *Summa Technologiae*, Wydawnictwo Literackie, Kraków, Poland, 2nd edition, 1966.

[2] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, et al., "Brain-computer interface technology: a review of the first international meeting," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 164–173, 2000.

[3] G. Pfurtscheller, C. Neuper, C. Guger, et al., "Current trends in Graz brain-computer interface (BCI) research," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 216–219, 2000.

[4] G. Pfurtscheller, "EEG event-related desynchronization (ERD) and event-related electro-encephalogram synchronization (ERS)," in *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*, E. Niedermayer and F. Lopes Da Silva, Eds., pp. 958–965, Williams & Wilkins, Baltimore, Md, USA, 4th edition, 1999.

[5] W. J. Williams, "Recent advances in time-frequency representations: Some theoretical foundations," in *Time Frequency and Wavelets in Biomedical Signal Processing*, M. Akay, Ed., IEEE Press Series in Biomedical Engineering, pp. 3–43, IEEE Press, Piscataway, NJ, USA, 1997.

[6] D. Harel, *Algorithmics: The Spirit of Computing*, Addison-Wesley, Reading, Mass, USA, 2nd edition, 1992.

[7] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[8] P. J. Durka and K. J. Blinowska, "Analysis of EEG transients by means of matching pursuit," *Annals of Biomedical Engineering*, vol. 23, no. 5, pp. 608–611, 1995.

[9] P. J. Durka, D. Ircha, C. Neuper, and G. Pfurtscheller, "Time-frequency microstructure of event-related electroencephalogram desynchronization and synchronization," *Medical & Biological Engineering & Computing*, vol. 39, no. 3, pp. 315–321, 2001.

[10] P. J. Durka, J. Żygierewicz, H. Klekowicz, J. Ginter, and K. J. Blinowska, "On the statistical significance of event-related EEG desynchronization and synchronization in the time-frequency plane," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1167–1175, 2004.

[11] M. B. Priestley, *Spectral Analysis and Time Series*, Academic Press, New York, NY, USA, 1981.

[12] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, New York, NY, USA, 1993.

[13] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, vol. 29, no. 4, pp. 1165–1188, 2001.

[14] P. J. Durka, D. Ircha, and K. J. Blinowska, "Stochastic time-frequency dictionaries for matching pursuit," *IEEE Trans. Signal Processing*, vol. 49, no. 3, pp. 507–510, 2001.

**Piotr J. Durka** received his M.S. and Ph.D. degrees in medical physics from Warsaw University, where he is currently an Assistant Professor. His research relates to the methodology of EEG analysis, mainly time-frequency signal processing. He introduced adaptive approximations (MP algorithm) to the EEG analysis; after a decade of successful applications, he aims at the unification of advanced signal processing and traditional, visual analysis of EEG.