

On the Unification in Electroencephalography

Piotr J. Durka*

Institute of Experimental Physics, Warsaw University, ul. Hoża 69 00-681 Warszawa

Electroencephalographic recordings (EEG) have been the main source of information on brains functional state for 70 years. Introduction of advanced signal processing methods in recent decades increased the fragmentation of this field and introduced a methodological gap between the clinical and research applications. This paper discusses assumptions, needed to accept the following thesis:

Adaptive time-frequency approximations of signals unify most of the univariate computational approaches to EEG analysis, and offer compatibility with its traditional (visual) analysis, used in clinical applications.

I. INTRODUCTION

A. Historical background

“Animal electricity” has been subject to scientific research since the end of the 18th century, when Galvani and Volta performed their famous experiments [1]. Electrical activity of the brain was first mentioned in 1875, in a grant report by R. Caton [2]. In 1929 the first electroencephalogram (EEG) was recorded from the surface of human skull by Hans Berger [3].

Year 1935 witnessed birth of the major fields of today’s clinical electroencephalography. F. Gibbs and H. Davis [4] showed association of 3/sec spike-wave complexes in EEG with epileptic petit mal absences, and A. L. Loomis et al [5] methodically studied human sleep EEG patterns and the stages of sleep. Also in 1935, the first electroencephalograph (Grass Model I) started the era of contemporary EEG recording: galvanometers, used in earlier decades to record EEG traces on photographic paper, were replaced by 3-channels preamplifier, and the recording was drawn by ink writer on rolls of paper. These rolls were later replaced by folded paper, and, currently, by digital storage and display of EEG traces. Also, contemporary amplifiers provide higher sensitivity and number of channels, but all these changes are quantitative rather than qualitative.

Finally, by the end of forties, Dawson [6] recorded first evoked potentials. He constructed an advanced mechano-electrical (analog) device for averaging brains potentials triggered by a stimulus [7]. Averaging was indispensable to show the event-related activity, which is normally invisible in the on-going EEG background.

These techniques, combined with *visual* analysis of recorded EEG traces, constitute the canon of contemporary clinical electroencephalography.

*URL: <http://durka.info>; Electronic address: durka@fuw.edu.pl

B. Computational techniques of EEG analysis

First attempts were made by Hans Berger in 1932 [8]; he was assisted by the physicist Dietsch, who applied Fourier analysis to short EEG sections. Then, quoting [9]: *The 1950s saw the early generation of automatic frequency analyzers approaching and eventually saw the end of these magnificent but mostly unused machines.*

Digital storage of the EEG time series opened unlimited possibilities of offline analysis, and triggered significant efforts towards algorithmic solutions. Studies published in last decades cover the whole spectrum of possible signal processing methods. This is partially due to the fact, that “first” application of given signal processing method to an old experimental paradigm or dataset usually fulfills the requirement of scientific novelty, and justifies publication.

C. The need for a unification

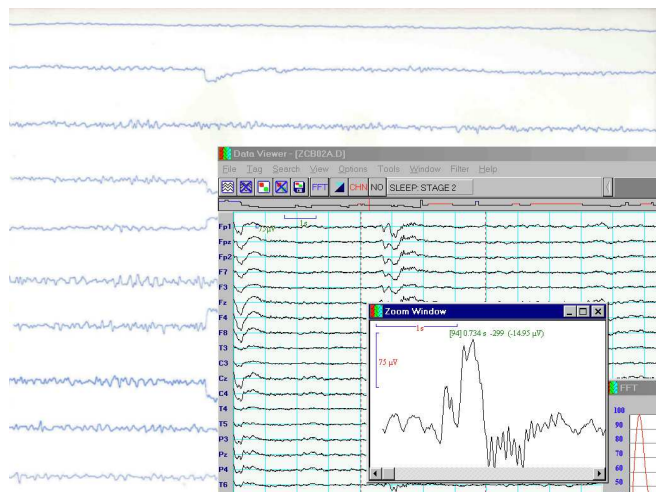


FIG. 1: 70 years of progress in clinical electroencephalography: from visual analysis of EEG traces on paper (background picture) to visual analysis of EEG traces displayed on CRT (front). Figure and caption reprinted from [10], © 2001 IEEE.

As mentioned in section I A, nowadays technology of

EEG recording is quite satisfactory: we can record simultaneously more channels than the number of electrodes which would fit on a skull, with sampling (in both time and amplitude) exceeding the assumed information content of the signal. Unresolved issues relate to the analysis of these huge amounts of data.

One of the major problems in biomedical sciences is the inter-subject variability. Especially in neurosciences, it may dramatically exceed the effects attributed to the investigated phenomena. This naturally leads to an unrealistic postulate that, to draw significant conclusions, each new hypothesis or method must be evaluated on a sufficiently large dataset, that is involving enough (probably at least thousands) human subjects. Otherwise, a coherent progress may be achieved only via comparison of different studies, performed and analyzed in a compatible way—to allow drawing common and coherent conclusions.

The only paradigm, up to now applied to a significant amount of cases, is the classic, visual analysis of EEG, involved in clinical and research applications. *Classic EEG evaluation always involved measuring frequency and/or amplitude with the help of simple rulers* [11]. Some of the EEG structures are even defined using terms like “cycles per second” and “microvolts” [12]. Nevertheless, new methods proposed for EEG analysis are mostly incompatible with such description. Therefore, they cannot be directly related to this indispensable knowledge base. Also, results offered by most of the advanced methods are seldom compatible with each other, so they do not seem to form a new, coherent knowledge base.

II. UNIVARIATE EEG ANALYSIS

This section briefly recalls the univariate linear and quadratic transforms, which were used in a significant amount of EEG studies.

A. Linear expansions

To explore the properties of an unknown signal f —in the absence of a quantitative model of its generation—we may use a linear expansion in a set of known functions $\{g_1(t), g_2(t), \dots, g_N(t)\}$:

$$f = \sum_{i=1}^N a_i g_i \quad (1)$$

If $\{g_i\}$ form an orthonormal basis, a_i are given simply by

$$a_i = \langle g_i, f \rangle = \int f(t) \overline{g_i(t)} dt \quad (2)$$

Traditionally, orthogonal $\{g_i\}$ sets were limited to Dirac’s deltas δ_i [31] and complex exponentials $e^{i\omega t}$, corresponding to the time-domain and spectral analyses. In recent

decades, new orthogonal bases were discovered, like e.g. wavelet packets or local cosines. The first and most popular orthogonal wavelet bases [32] are built from dilations and translations of a single function $\psi(\frac{t-u}{2^j})$ [33].

To interpret (1), we usually relate those g_i , that correspond to the largest $|a_i|$, to some properties of f . For example, large value of $\langle f, e^{i\omega_0 t} \rangle$ (spectral peak at ω_0) suggests, that f contains periodic activity with base frequency near ω_0 . However, presence of the same periodic activity will not be so evident if $|a_i|$ are calculated e.g. in a wavelet basis, since in this case it will be diluted across many wavelets.

B. Time-frequency distributions

Increasing need to study the detailed picture of event-related changes of energy—e.g. in the event-related EEG (de)synchronization studies [13] or research on brain-computer interfaces [14]—turns our attention to the quadratic estimates of the time-frequency energy density. The most natural (at least for a physicist) is the Wigner–Ville transform \mathcal{W} :

$$\mathcal{W}f = \int f\left(t + \frac{\tau}{2}\right) \overline{f\left(t - \frac{\tau}{2}\right)} e^{-i\omega\tau} d\tau \quad (3)$$

However, it suffers from severe interferences (cross-terms [34]), so several reduced interference distributions (RID) were proposed to minimize this drawback. It is achieved via the smoothing kernel ϕ , which decreases the influence of cross-terms at a cost of lower time-frequency resolution:

$$\mathcal{C}f = \int \int \int \phi(\theta, \tau) e^{-i(\theta\tau + \theta t - \omega u)} \cdot f\left(t + \frac{\tau}{2}\right) \overline{f\left(t - \frac{\tau}{2}\right)} d\tau du d\theta. \quad (4)$$

The most popular time-frequency distributions are spectrogram (squared modulus of the short-time Fourier transform) and scalogram (squared modulus of continuous wavelet transform). All of them can be derived from (4) with different kernels ϕ , and they give better or worse results for different types of signals. Therefore, it is possible to obtain a clear time-frequency picture of a given signal f , if we find an optimal kernel ϕ for its particular content. However, there is no general recipe for this choice. Also, these quadratic representations do not offer the direct parameterization of signal’s features such as (1).

III. ADAPTIVE TIME-FREQUENCY APPROXIMATION

In the ideal situation we would like to have a linear expansion like (1), but with functions g_i representing *all* the relevant structures present in f . Mere expansion of

the set $\{g_i\}$ beyond a basis does not solve the problem. For the representation of each signal f we must *choose* the set functions $\{g_{\gamma_i}\}_{i=1..M}$ adapted to represent its particular content. As a result, we get an approximation [35]:

$$f \approx \sum_{i=1}^M a_i g_{\gamma_i} \quad (5)$$

Criterion of choice can be based upon the minimal norm of the residue. We can define an optimal M -approximation as an expansion, minimizing the following error ϵ :

$$\epsilon = \left\| f - \sum_{i=1}^M w_i g_{\gamma_i} \right\| \quad (6)$$

Finding such an optimal approximation is NP-hard (intractable), but a suboptimal solution can be found by means of an iterative procedure like the matching pursuit algorithm (MP), proposed by Mallat and Zhang in 1993 [15]. However, even this sub-optimal solution is computationally intensive, so the first practical applications were not possible before the last decade.

A. Matching pursuit algorithm

In the first step of MP, the waveform g_{γ_0} which best matches the signal f is chosen from the dictionary D . In each of the consecutive steps, the waveform g_{γ_n} is matched to the signal $R^n f$, which is the residual left after subtracting results of previous iterations:

$$\begin{cases} R^0 f = f \\ R^n f = \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n} + R^{n+1} f \\ g_{\gamma_n} = \arg \max_{g_{\gamma_i} \in D} |\langle R^n f, g_{\gamma_i} \rangle| \end{cases} \quad (7)$$

Orthogonality of $R^{n+1} f$ and g_{γ_n} in each step implies energy conservation

$$\|f\|^2 = \sum_{n=0}^{m-1} |\langle R^n f, g_{\gamma_n} \rangle|^2 + \|R^m f\|^2 \quad (8)$$

For a complete dictionary the procedure converges to f

$$f = \sum_{n=0}^{\infty} \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n} \quad (9)$$

B. Time-frequency dictionaries of Gabor functions

Time-frequency dictionaries are composed—apart from the complete Dirac and Fourier bases—from the Gabor functions, since these functions provide optimal joint time-frequency localization [16]. Real valued Gabor can be expressed as

$$g_{\gamma}(t) = K(\gamma) e^{-\pi \left(\frac{t-u}{s}\right)^2} \cos(\omega(t-u) + \phi), \quad (10)$$

where $K(\gamma)$ is such that $\|g_{\gamma}\| = 1$. Parameters $\gamma = \{u, \omega, s\}$ of the possible Gabor functions constitute a 3-dimensional [36] continuous space, from which a finite dictionary must be chosen for an implementation of the procedure (7). Originally (in [15]), dictionary's parameters were chosen from predefined dyadic sequences. However, any fixed scheme of subsampling the space of possible dictionary's functions leads to a statistical bias of the resulting decompositions. A solution proposed in [17] relies on stochastic dictionaries, in which parameters $\{u, \omega, s\}$ are drawn from uniform distributions across ranges defined by sizes of the signal and the dictionary.

C. Estimate of signal's energy density

If we find the signal's approximation (5), we can also solve the problem of cross-terms present in the distributions from section II B. Namely, calculating the Wigner distribution of expansion (5) would yield

$$\begin{aligned} \mathcal{W}f &\approx \mathcal{W} \left(\sum_{i=1}^M a_i g_{\gamma_i} \right) = \\ &= \sum_{i=1}^M a_i^2 \mathcal{W}g_{\gamma_i} + \sum_{i=1}^M \sum_{j \neq i}^M a_i \bar{a}_j \mathcal{W}(g_{\gamma_i}, g_{\gamma_j}) \end{aligned} \quad (11)$$

where $\mathcal{W}(g_{\gamma_i}, g_{\gamma_j})$ is a cross-Wigner transform of g_{γ_i} and g_{γ_j} given by

$$\mathcal{W}(g_{\gamma_i}, g_{\gamma_j}) = \int g_{\gamma_i} \left(t + \frac{\tau}{2} \right) \overline{g_{\gamma_j} \left(t - \frac{\tau}{2} \right)} e^{-i\omega\tau} d\tau \quad (12)$$

Double sum in (11) contains all the cross-terms. Owing to the representation (5), we can omit them explicitly and construct the time-frequency representation of signal's energy density from the first sum, containing auto-terms:

$$\mathcal{E}f = \sum_{i=1}^M a_i^2 \mathcal{W}g_{\gamma_i} \quad (13)$$

Energy conservation of this distribution is easily demonstrated (c.f. [15]).

IV. UNIFICATION

In this sections we collect and discuss the statements, needed for the acceptance of the thesis formulated in the Abstract:

Adaptive time-frequency approximations of signals unify most of the univariate computational approaches to EEG analysis, and offer compatibility with its traditional (visual) analysis, used in clinical applications.

Lemma 1 *Matching pursuit sub-optimal solution to the problem of adaptive approximation is suitable for EEG analysis.*

Sub-optimal solution of an intractable problem (section III) must have its price. Mathematical examples of failures in pattern recognition, due to the greedy strategy (7) applied by the matching pursuit, were presented e.g. in [18, 19]. These cases did not relate to structures encountered in biomedical signals, but the lack of a relevant counterexample does not deny its existence. Figure 2 gives an example of a case referring directly to the transient oscillatory activity of the kind which may be actually present in the EEG.

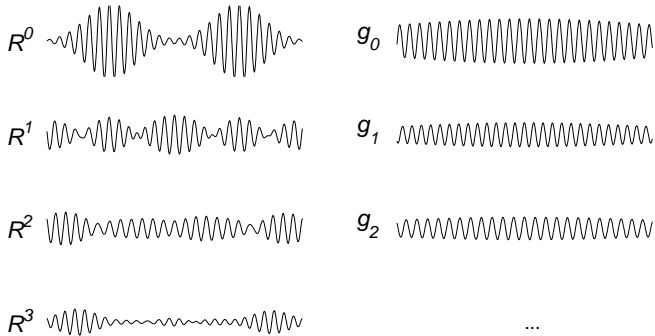


FIG. 2: A failure in feature extraction: R^0 —analyzed signal, g_0 —function fitted in the first iteration by the MP algorithm. Horizontal—time, vertical—amplitude, both in arbitrary units (simulated signals).

Signal (R^0) in Figure 2 is composed from two Gabor functions, both of them actually present in the dictionary D used for the decomposition (Section IIIB). In spite of that, we observe (in the right column) that the first function (g_0) fitted to the signal is completely different from either of the two functions, from which the signal was composed! According to (7), MP algorithm has chosen function g_i giving the largest product $\langle f, g_i \rangle$ in a *single* step. Of course, taking into account the next steps, this decision was definitely not optimal. Choosing the two Gabors, which were exactly represented in D , would explain 100% of signal’s energy in only 2 iterations, contrary to the residues produced in the left column of Figure 2 as a consequence of the first choice.

However, such an effect occurs only if both Gabors present in the signal have not only the same frequencies, but also *exactly* the same phase. Such a “coincidence” would be possible in a biological signal *only* if both the structures were produced by the same generator. And still, MP represents them jointly only if their time centers are close enough. Their larger displacement would result in separate representation even in such a synchronized case. Therefore, we may argue that this effect, mathematically classified as a failure of the sub-optimal procedure, is actually a welcome feature in the analysis of physiological signals. It was explored in [20] for de-

tecting the synchronized spiking activity, as opposed to series of unrelated EEG spikes.

Even if we were able to calculate the optimal adaptive decomposition minimizing error (6), it would have one more disadvantage: namely, the set of M functions g_{γ_i} chosen to optimally represent signal f , may significantly differ from the optimal set of $M + 1$ functions chosen for the same signal f (from the same dictionary D). With the iterative MP solution (7), $\{g_{\gamma_i}\}_{i=1..M}$ will be the same in both decompositions.

Considering that the features related to the sub-optimality of the MP solution turn to be advantages rather than drawbacks in the analysis of biomedical signals, we conclude that matching pursuit algorithm is the correct implementation for an adaptive time-frequency approximation of EEG.

Lemma 2 *Matching pursuit decomposition (5) is asymptotically well defined and does not depend on arbitrary settings.*

Calculating representation (5) of a signal f via matching pursuit in given dictionary D is a deterministic procedure, given by (7). Therefore, for a given signal, representation (5) depends only on the dictionary D . For the Gabor dictionaries discussed in section IIIB, the only free parameter is their size. It influences MP decomposition in a predictable way: larger dictionaries improve the quality of the decomposition, that is less waveforms are needed to achieve the same error ϵ in (6). However, starting from a certain density of dictionary’s waveforms in the space of their parameters $\{u, \omega, s\}$, this improvement reaches asymptotically a saturation. At this point, which we shall tentatively term a “sufficient density of a Gabor dictionary”, decompositions in larger dictionaries, or in different stochastic realizations of the same dictionary, are indistinguishable (of course this relates only to the structures coherent with the dictionary, c.f. [15]).

This qualitative reasoning—although still not backed up by a quantitative proof—is well supported by numerical experiments and years of real-world applications. Problems with mathematical derivations stem from the non-linearity of the MP procedure (7) and variable content of the analyzed signals. Therefore, the notion of “sufficient density” of a Gabor dictionary is up to now defined only empirically. Nevertheless, for the following we shall use it as a requirement for (5). For reasonable dimensions of signals, such densities are achieved with the dictionary sizes easily implementable on today’s computers.

Conjecture 3 *Relevant content of the EEG signal is coherent with Gabor time-frequency dictionaries.*

This statement depends on the meaning of “relevant EEG content”, which, unfortunately, does not have a strict definition.

EEG originates from postsynaptic potentials of firing neurons; only if a large ensemble is working synchronously, the signal can reach a level measurable from

the skull electrodes. A single electrode reads a potential from an arbitrary number of such ensembles, spatially filtered by different conductivities of cortex, dura mater, skull and scalp. Recorded EEG is not only affected by possibly non-linear interactions between these ensembles, but the signal may also contain extra-cerebral—biological or not—potentials (artifacts).

Rhythmic activity has been recognized as a prominent feature of the signal since the beginnings of EEG recordings. It appears as transient—rather than permanent—oscillations. These oscillations are produced by large masses of neurons (typically $10^4 - 10^7$ [21]). Synchronization and desynchronization of these neural masses cannot be instantaneous, so the oscillations exhibit some waxing and waning. Gabor functions provide a concise general model for such signals.

“Relevant content of EEG” can be also translated to “relevant structures used to date in EEG analysis”. That would refer mainly to the visual analysis, discussed in lemma 6 and 7.

Finally, non-permanence of this statement—underlined by using “conjecture” rather than “lemma”—relates also to the fact that, if any other structures would appear to be important in this context, they can be added to the dictionaries without a negative impact on the convergence of the MP procedure (7).

Conjecture 4 *All the relevant EEG structures, which can be found via linear expansions in different bases, are included in the MP parameterization.*

As discussed in section II A, exploratory value of linear expansions relates to g_i with the largest weights in expansion (1)—only those are usually treated as reflecting the significant signal’s structures. Accepting Conjecture 3 we assume, that all relevant EEG structures are well represented in the Gabor dictionary. Also, waveforms used in linear expansions (1) can be more or less exactly assimilated by smooth functions (10) [37]. Therefore, we may expect that if a signal contains a mixture of oscillatory and transients activities, and each of them would be highlighted in a different expansion (1), then all of them will be efficiently represented in the MP approximation (5).

As an example, we may quote a study of pharmacological EEG [22], where traditional spectral estimators of spindling activity between 12 and 14 Hz were replaced by the energy of relevant structures estimated from MP expansion. It provided full compatibility with the traditional approach and better accordance with physiological expectations, owing to the increased sensitivity of MP-based estimator.

However, the correspondence to which we refer in this Conjecture, relates only to EEG *structures*, which should be parameterized by the MP decomposition as efficiently as in any of the linear expansions (1). But certain *statistical* properties of the signal can be derived only from its full representation e.g. in wavelet or Fourier bases. On the other hand, properties of MP expansion can be

used to construct completely new measures, like e.g. the Gabor atom density (GAD), proposed for prediction of epileptic seizures in [23].

Lemma 5 *Matching pursuit provides the most general solution to the problem of non-unique estimates of time-frequency energy density, applicable for EEG analysis.*

All the time-frequency distributions from section II B are estimates of the same magnitude, that is energy density of the signal. However, estimates derived from different approaches (e.g. Wigner, RID, CWT or spectrogram) differ significantly, mainly due to the variable contribution of cross-terms. Also, estimates obtained from the same approach may look quite different depending on their parameters, like e.g. the length of the time window in spectrogram, different wavelets in CWT or smoothing kernels ϕ in RID.

Estimate (13), derived from the MP decomposition, contains the auto-terms, corresponding to structures present in expansion (5), and no cross-terms (section III C). Therefore, for a signal coherent with the dictionary used in the MP decomposition, estimate (13) approximates the intersection of its different quadratic estimates of energy density. This common part corresponds to the generalized auto-terms, contrary to the cross-terms variable between estimates. Therefore, if we accept Conjecture 3, this lemma is also proved. Even if we doubt the coherence of EEG signal with the Gabor dictionary, this approach is currently the best candidate for a unifying approach, owing to the uniqueness discussed in Lemma 2, since all the other quadratic estimates depend on an arbitrary choice of parameters.

Recent studies of brains electrical activity, recorded from the skull [13, 24] and from the cortex [25], confirmed also that (13) provides resolution superior to the previously applied quadratic methods.

Lemma 6 *Among the contemporary methods of signal analysis, adaptive approximations with time-frequency dictionaries provide the best correspondence to visual EEG analysis.*

The knowledge, verified through decades of clinical applications, consists mainly of associations between appearances of rhythmic activities and certain waveforms in EEG with physiologically relevant states or pathologies, like e.g. level of alertness, sleep depth or epileptic seizure. These rhythms and waveforms constitute the dictionary of clinical EEG analysis. Due to a limited repeatability of their visual detection, the field is often considered an art rather than science [21]. Mapping this dictionary into mathematical terms will allow the application of repeatable scientific methods.

EEG rhythms ($\delta, \theta, \alpha, \beta, \gamma$) are approximately assigned to frequency bands. Therefore, if we assume that those frequency bands are fixed, we can detect their

presence from the relative power of the corresponding frequency band in the Fourier spectrum. However, power spectrum is a property of the whole analyzed epoch, and does not provide the information on the time extent of the given oscillatory phenomenon.

Appearance and disappearance of a rhythm can be clearly detected from a time-frequency map of signal's energy density, obtained from one of the quadratic methods discussed in section II B. However, stating that oscillations with frequency f last from the time t_1 to t_2 requires an analysis (post-processing) of a redundant map. Results will depend on the combination of the applied post-processing method and the chosen estimate of the energy density.

On the contrary, unequivocally defined (Lemma 2) MP expansion (5) gives *explicitly* the time span and frequency of detected oscillations.

Transient EEG structures—sleep spindles, K-complexes, epileptic spikes and many others—are obviously undetectable from the spectral estimates. They may exhibit certain, more or less typical, time-frequency signatures, but in the time-frequency maps of energy density the phase information is lost. We may imagine using the complex transforms (3) or (4) for their detection, but as previously mentioned we are confronted with continuum of possible transforms and post-processing algorithms. Adaptive approximations (5) provide today the only method of an explicit description of a significant variety of waveforms. As presented in [20], this description is efficient also for non-oscillating structures like epileptic EEG spikes.

Conjecture 7 *The correspondence from Lemma 6 can bridge the art of visual EEG analysis and scientific methods of signal processing.*

Adaptive time-frequency approximation is currently the only signal processing method, offering description of structures, present in a signal, explicitly in terms of their physical parameters like amplitude, frequency or time width. Similarly to the way in which an expert evaluates EEG, the algorithm concentrates on most prominent, local structures, rather than general properties of the whole analyzed epoch.

Visual analysis of EEG relies primarily on detecting occurrences of rhythms and other waveforms, called graphoelements. In some cases, their classic definitions are given already in time-frequency terms. For example, sleep spindles are defined (c.f. [12]) as structures of least 0.5 sec duration, frequency between 12 and 14 Hz and amplitude above 15–25 μV . Such a definition can be directly applied to filter the decomposition (5) in the space of the parameters of functions g_{γ_i} , fitted to the analyzed

signal. Waveforms g_{γ_i} , conforming to such ranges of parameters, represent sleep spindles—as verified by comparison with their visual detection in [26].

However, not all the graphoelements have such strict, mathematical definitions. The very name “graphoelement” (rather than waveforms) suggests, that their visual detection may rely on different aspects of its shape. Nevertheless, most of these shapes can be effectively approximated by Gabor functions. In such cases we may a posteriori adapt the ranges of the parameters of g_{γ_i} , to maximize the concordance with visual detection. This resembles the classical paradigm of expert systems, however, using the MP parameterization (5) has the major advantage of operating in the space of physically meaningful and universal parameters, like time width, amplitude or frequency. An example of a successful implementation of such an approach and its broader discussion can be found in [20].

Finally, in relation to the clinical analysis, we may quote preliminary results of a recent study [27] performed on a relatively large material (21 overnight EEG recordings) and related to a standard clinical task:

Determination of the sleep depth (stages assigned to 20-sec epochs) across the whole night is a part of the clinical routine. In attempts to automatize this tedious and important task, greatest difficulties were encountered in determination of the deep sleep stages (III and IV), in spite of their relatively simple definition. It is based upon the relative contribution of the slow wave activity (SWA), which is defined as waves of frequency below 2 Hz and peak-to-peak amplitude above 75 μV . In stage IV, more than 50% of epoch's time is occupied by SWA. If SWA occupies between 20% and 50% of the epoch's time, it's said to represent sleep stage III. MP expansion (5) is the only one (in contemporary signal processing), which offers explicitly the time span of detected structures (it's proportional to s from Eq. 10). Detection of deep sleep stages, based upon this feature, yielded concordance with the traditional staging on a level similar to the inter-expert agreement, evaluated in [28].

V. DISCUSSION

To gain a proper perspective on the proposition of accepting a non-conclusive argumentation from the previous section, let us consider another theorem, which is generally *believed* to be true, and, although still not proved, is used by most of us on a daily basis (it was quoted also in this paper). Namely, according to the section III, the problem of finding an optimal approximation (5) is NP-hard. It means that its computational complexity is growing with the size of the problem (in this case dimension of the signal) exponentially—in computer science this term is used as synonym of “faster than any polynomial”. Such problems, even for moderately-sized input data, require times easily exceeding e.g. the age of the Universe.

We can formulate an almost equivalent decision problem, asking whether for given D , ϵ_1 and f exists an M -approximation giving error (6) smaller than ϵ_1 . It can be proved to belong to the NP-complete class [29], that is problems possessing the two following properties [30]:

- Finding a solution requires exponential (non-polynomial time), but given solution can be checked in polynomial (P) time [38]
- Solution of each of the problem from the NP-class can be translated into solution of any other problem from this class in a polynomial time.

It is generally believed, that NP-complete problems do not have polynomial-time solutions, yet this basic theorem ($P \neq NP$) is still not proved [39]. Factorization of large numbers also belongs to the NP-complete class. Strength of the public key cryptography depends on factorization. Nevertheless, we are still using public key cryptography in transmission of sensitive data (e.g. credit card numbers, bank transfers) over Internet. Our belief is based upon the lack of counterexamples and ex-

perimental evidence given by years of application and up-to-date security offered by these methods. Similarly, we may provisionally accept the main thesis of this paper.

VI. CONCLUSION

Owing to the recent advances in mathematics and computer technology, we can achieve a unified, adaptive and high-resolution description of the EEG time series in terms of well-defined waveforms. These waveforms can be related to the graphoelements and rhythms, used for decades in visual EEG analysis.

Application of this paradigm will have a tremendous impact on the coherence of the EEG research and automatization of some of the tedious and expensive procedures, necessary in clinical EEG practice. To facilitate its practical implementation, software for matching pursuit decomposition of signals is freely available at <http://eeg.pl/mp>.

-
- [1] A. N. Martonosi, *Acta Biochimica Polonica* **47**, 493 (2000).
- [2] M. A. B. Brazier, *A History of the Electrical Activity of the Brain, The First Half-Century* (Pitman Medical Publishing, 39 Parker Street, London W.C. 2, 1961).
- [3] A. Beck, *Centerblatt für Physiologie* **4**, 572 (1890).
- [4] F. A. Gibbs and H. Davis, *Am. J. Physiol* **113**, 49 (1935).
- [5] A. L. Loomis, E. N. Harvey, and G. A. Hobart, *Science* **82**, 198 (1935).
- [6] G. D. Dawson, *J Neurol Neurosurg Psychiatry* **10**, 134 (1947).
- [7] G. D. Dawson, *Electroencephalogr and clinical neurophysiology* **6**, 65 (1954).
- [8] G. Dietsch, *Pflügers Arch. Ges. Physiol.* **230**, 106 (1932).
- [9] E. Niedermayer, in *Electroencephalography: Basic Principles, Clinical Applications and Related Fields* (Williams & Wilkins, 1999), pp. 1–14, 4th ed.
- [10] P. J. Durka and K. J. Blinowska, *IEEE Engineering in Medicine and Biology Magazine* **20**, 47 (2001).
- [11] F. Lopes Da Silva, in *Electroencephalography: Basic Principles, Clinical Applications and Related Fields* (Williams & Wilkins, 1999), pp. 1135–1163, 4th ed.
- [12] A. Rechtschaffen and A. Kales, eds., *A manual of standardized terminology, techniques and scoring system for sleep stages in human subjects*, no. 204 in National Institutes of Health Publications (US Government Printing Office, Washington DC, 1968).
- [13] P. J. Durka, D. Ircha, C. Neuper, and G. Pfurtscheller, *Med Biol Eng Comput* **39**, 315 (2001).
- [14] P. Durka, *EURASIP Journal on Applied Signal Processing Trends in Brain-Computer Interfaces* (2005), submitted.
- [15] S. Mallat and Z. Zhang, *IEEE Transactions on Signal Processing* **41**, 3397 (1993).
- [16] S. Mallat, *A wavelet tour of signal processing* (Academic Press, New York, 1999), 2nd ed.
- [17] P. J. Durka, D. Ircha, and K. J. Blinowska, *IEEE Trans Signal Process* **49**, 507 (2001).
- [18] R. A. DeVore and V. N. Temlyakov, *Advances in Computational Mathematics* **5**, 173 (1996).
- [19] S. S. Chen, D. L. Donoho, and M. A. Saunders, *SIAM Review* **43**, 129 (2001).
- [20] P. Durka, *Physical Review E* (2004), (accepted).
- [21] E. Niedermayer and F. Lopes Da Silva, *Electroencephalography: Basic Principles, Clinical Applications and Related Fields* (Williams & Wilkins, 1999), 4th ed., p. 1154.
- [22] P. Durka, W. Szelenberger, K. Blinowska, W. Androsiuk, and M. Myszka, *Journal of Neuroscience Methods* **117**, 65 (2002).
- [23] C. C. Jouny, P. J. Franaszczuk, and G. K. Bergey, *Clin. Neurophysiol.* **114**, 426 (2003).
- [24] P. Durka, J. Żygierewicz, H. Klekowicz, J. Ginter, and K. Blinowska, *IEEE Transactions on Biomedical Engineering* (2004), (in print).
- [25] P. J. Franaszczuk, G. Bergey, P. J. Durka, and H. Eisenberg, *Electroencephalogr Clin Neurophysiol* **106**, 513 (1998).
- [26] J. Żygierewicz, K. J. Blinowska, P. J. Durka, W. Szelenberger, S. Niemcewicz, and W. Androsiuk, *Clin Neurophysiol* **110**, 2136 (1999).
- [27] U. Malinowska, Master's thesis, Warsaw University (2004).
- [28] R. Norman, I. Pal, C. Stewart, J. Walsleben, and D. Rapoport, *Sleep* **23**, 901 (2000).
- [29] G. Davis, Ph.D. thesis, New York University (1994), <ftp://cs.nyu.edu/pub/wave/report/DissertationG-Davis.ps.Z>.
- [30] D. Harel, *Algorithmic: The Spirit of Computing* (Addison-Wesley, 1992), 2nd ed.

- [31] Discrete Dirac's delta is a sequence of zeros except for unity in the point i , that is $\delta_i(k) = 1$ for $k = i$ and zero otherwise.
- [32] Of course only very special choices of ψ give an orthogonal basis, c.f. [16]. Actually, it was not before 1980. that first such orthogonal sets were discovered.
- [33] This implies the characteristic "zooming property" of wavelets, that is increased time resolution and decreased frequency resolution at high frequencies. However, in the EEG analysis this seems to be more of a drawback than advantage.
- [34] Cross-terms are the area of a time-frequency energy density estimate, which may be interpreted as indicating false presence of signal's activity in given time and frequency coordinates. Their origin can be deduced e.g. from equation (11).
- [35] Approximation appears instead of the exact expansion, because to choose g_i from a set large enough to represent efficiently relevant signal's structures, we must give up the orthogonality. Adaptively chosen orthogonal representations, available in terms of wavelet packets, do not provide enough flexibility
- [36] Phase ϕ is optimized separately in practical implementations
- [37] E.g. some wavelets contain discontinuities, but in the context of spatially-filtered EEG (c.f. Lemma 3) smooth Gabor functions should be more adequate for its description
- [38] The name "nondeterministic-polynomial" stems from the lack of a deterministic algorithm finding this answer in polynomial time ("guessing" the right answer is non-deterministic).
- [39] It may be also undecidable from the currently applied axioms.