Sparse Correlation Kernel Analysis and Evolutionary Algorithm-based Modeling of the Sensory Activity within the Rat's Barrel Cortex

Mariofanna Milanova[1], Tomasz G. Smolinski[2], Grzegorz M. Boratyn[2], Jacek M. Zurada[2], and Andrzej Wrobel[3]

[1] Department of Computer Science, University of Arkansas at Little Rock, Little Rock, AR 72204, USA
mgmilanova@ualr.edu
http://pandora.compsci.ualr.edu/
[2] Computational Intelligence Laboratory, Department of Electrical and Computer Engineering, University of Louisville, Louisville, KY 40292, USA
{tomasz.smolinski, grzegorz.boratyn, jacek.zurada}@louisville.edu
http://www.ci.louisville.edu/
[3] Laboratory of Visual System, Department of Neurophysiology, Nencki Institute of Experimental Biology, 3 Pasteur Street, 02-093 Warsaw, Poland
wrobel@nencki.gov.pl
http://www.nencki.gov.pl/labs/vslab/vislab.html

Abstract. This paper presents a new paradigm for signal decomposition and reconstruction that is based on the selection of a sparse set of basis functions. Based on recently reported results, we note that this framework is equivalent to approximating the signal using Support Vector Machines. Two different algorithms of modeling sensory activity within the barrel cortex of a rat are presented. First, a slightly modified approach to the Independent Component Analysis (ICA) algorithm and its application to the investigation of Evoked Potentials (EP), and second, an Evolutionary Algorithm (EA) for learning an overcomplete basis of the EP components by viewing it as probabilistic model of the observed data. The results of the experiments conducted using these two approaches as well as a discussion concerning a possible utilization of those results are also provided.

1 Introduction

The standard methods for decomposition and analysis of evoked potentials are band pass filtering, identification of peak amplitudes and latencies, Principal Component Analysis (PCA) and wavelet-based analysis. A common way to represent real-valued signals is using a linear superposition of basis functions. One might roughly characterize the second-order methods (PCA and Factor Analysis) by saying that their purpose is to find a faithful representation of the

data in the sense of signal's reconstruction (e.g. based on the mean-square error measure) [1], [2]. In contrast, the highest-order methods (Projection Pursuit, Blind De-convolution, Independent Component Analysis) are characterized by their attempt of finding a meaningful representation of the signal. Of course, meaningfulness is a task-dependant property [3], [4].

The standard wavelet analysis gives the coefficients for expressing the signal as a linear combination of "wavelet packets", which can include scaled and translated versions of the father and mother wavelets, as well as scaled and translated versions that contain additional oscillations. The wavelet coefficient is an inner product of the wavelet and the data. Bases such as the Fourier or wavelet can provide a useful representation of some signals, but they are limited, because they are not specialized for the signals under consideration [5].

An alternative and more general method of signal representation uses the so-called "overcomplete bases" (also called overcomplete dictionaries), which allows for a greater number of basis functions than samples in the input signals [6], [7], [8]. Overcomplete bases are typically constructed by merging a set of complete bases (e.g. Fourier, wavelet, Gabor) or by adding basis functions to a complete basis (e.g. adding frequencies to a Fourier basis).

Relatively little research has been done in the area of decomposition of Evoked Potentials (EP) using non-orthogonal components. Olshausen and Field [9], [10] pretend that the basis functions shared many properties with neurons in primary visual cortex, suggesting that overcomplete representations might be a useful model (see also [11]). Under an overcomplete basis, the decomposition of a signal is not unique, but this can be in our advantage – we have greater flexibility in capturing structure in the data.

Subsequently, two different algorithms of modeling sensory activity within the barrel cortex of a rat are presented. First, we slightly modify the traditional approach to the Independent Component Analysis (ICA) and apply it to EP's investigation, and second, we propose a new, evolutionary algorithm-based approach to decomposition of evoked potentials. More specifically, we propose an algorithm for learning an overcomplete basis of the EP components by viewing it as probabilistic model of the observed data. In an overcomplete basis, the number of basis vectors is greater than the dimensionality of the input data. Overcomplete representation has been advocated because it has greater robustness in presence of noise, it is more sparse, and has greater flexibility in matching structure in the data [9]. From this model, we derive a simple, robust learning algorithm by maximizing the data likelihood over the modeled data based on the basis functions.

After surveying published work in the literature within this field, it appears that our approach of decomposition of EP's in terms of a set of overcomplete basis functions and the process of learning them by evolutionary algorithm is unique.

The paper is organized into the following sections: in Sect. 2, we discuss the proposed data model an then, in Sect. 3, the formal learning algorithm is defined as well as our evolutionary algorithm-based methodology used to learn the set of basis functions. In Sect. 4, we present results of decomposition and comparison between the proposed methods.

2 Bayesian Motivated Model

The primary step in measuring the form and variance of evoked potentials (EP) or event-related potentials (ERP) is to decompose them into parts. Each part has three aspects. The first one is the elementary curve. The elementary curves are called basis functions. When a selected set of such curves is added together, the sum should closely conform to the shape of the ERP. Second, each part has a set of numbers or coefficients that denote its amplitude. Third, each part defines one or more coordinates, and the set of numbers denotes distances along the axes of a coordinate system. The number of coefficients in the set of basis functions specifies the dimensions of the measurement space. From this point of view the measurement of each ERP gives a vector in that space [12].

We assume that each data vector can be described with a set of basis functions plus some additive noise:

$$\mathbf{x} = \mathbf{M}\mathbf{a} + \varepsilon, \tag{1}$$

where $\mathbf{x}$ is the signal, $\mathbf{M}$ is the matrix of basis functions, $\mathbf{a}$ is the vector of coefficients (i.e. the representation of the signal), and ε represents Gaussian noise.

Let n denote the number of time points in each recorded EP. Let $x_i(t)$ denote the value of each i-th recorded EP at time t, where $t = 1,2,\ldots,n$, and $i = 1,2,\ldots,m$. The model specifies that $x_i(t)$ is a weighted sum of unknown EP components, where the weights depend on the EP component and each individual measurement. Let r denote the number of EP components, which is assumed known, and let $M_j(t)$ denote the unknown value of EP component (basis function) j at time t, where $j = 1,2\ldots,r$. Let a_{ij} be the unknown weight of EP component j for individual signal i. Then we assume a model of the form:

$$x_i(t) = \sum_{j=1}^{r} a_{ij} M_j(t) + \varepsilon_i(t), \tag{2}$$

where $\varepsilon_i(t)$ represents Gaussian noise.

The unknown parameters to be estimated are a_{ij} and $M_j(t)$. Developing efficient algorithms to solve this equation is an active research area. A given data point can have many possible representations, nevertheless this redundancy can be removed by a proper choice for the prior probability of the basis coefficients:

$$P(a_{ij}) = \exp(-S(a_{ij})), \tag{3}$$

where:

$$S(a_{ij}) = \beta \log\left(1 + \left(\frac{a_{ij}}{\sigma}\right)^2\right), \tag{4}$$

and β and σ are scaling factors.

This specifies the probability of the alternative representations. Standard approaches to signal representation do not specify a prior for the coefficients. Assuming zero noise, representation of a signal in this model is unique. If $\mathbf{M}$ is invertible, the decomposition of the signal $\mathbf{x}$ is given by $\mathbf{a} = \mathbf{M^{-1}x}$. Since $\mathbf{M^{-1}}$ is expensive to compute, the standard models use the basis matrices that are easily inverted, by, for instance, restricting the basis functions to be orthogonal (in PCA) or by limiting the set of basis functions to those, for which there exist fast computational algorithms, such as Fourier or wavelet analysis. Usually, to define a unique set of basis functions we have to impose unrealistic mathematical constraints. For example, PCA assumes that the data distribution has a Gaussian structure and fits in appropriate orthogonal basis. ICA, generalized PCA, assumes that the coefficients have non-Gaussian structure and this allows the basis functions to be non-orthogonal. In all of these techniques, the number of basis vectors is equal to the number of inputs.

A more general approach is to use the information theory and the probabilistic formulation of the problem [13]. Rather than making prior assumption about the shape or form of the basis functions, those functions are adapted to the data using an algorithm that maximizes the log-probability of the data under the model.

The coefficients from (1) can be inferred from $\mathbf{x}$ by maximizing the conditional probability of a given $\mathbf{x}$, $P(\mathbf{a}|\mathbf{x}, \mathbf{M})$, which can be expressed via Bayes' rule as:

$$\mathbf{a} = \arg\max_{\mathbf{a}} P(\mathbf{a}|\mathbf{x}, \mathbf{M}) \propto \arg\max_{\mathbf{a}} P(\mathbf{x}|\mathbf{a}, \mathbf{M}) P(\mathbf{a}). \tag{5}$$

The first term of the right hand side of the proportion specifies the likelihood of the signal under the model for a given state of the coefficients:

$$P(\mathbf{x}|\mathbf{a}, \mathbf{M}) \propto \exp\left(-\frac{\lambda}{Z_{\sigma N}} |\mathbf{x} - \mathbf{Ma}|^2\right), \tag{6}$$

where $Z_{\sigma N}$ is normalizing constant, $\lambda = 1/\sigma^2$, and σ is the standard deviation of the additive noise. The second term specifies the prior probability distribution over the basis coefficients, where:

$$P(\mathbf{a}_i) = \prod_j P\left(a_{ij}\right). \tag{7}$$

Thus, the maximization of the log-probability in (5) becomes:

$$a_{ij} = \arg\min_{a_{ij}} \left(\frac{\lambda_N}{2} \sum_{t=1}^{n} \left| x_i(t) - \sum_{j=1}^{r} a_{ij} M_j(t) \right|^2 + \sum_{j=1}^{r} S(a_{ij}) \right). \tag{8}$$

This formulates the problem as one of the density estimation and is equivalent to minimizing the Kullback-Leibler (KL) divergence between the model density and the distribution of the data.

The functional (8) that is minimized, consists of an error term and a sparseness term.

Based on recently reported results [14], we note that this framework is equivalent to approximating the signal using Support Vector Machines (SVM).

In SVM we approximate the signal $x(t)$ as:

$$M_i(t) = K(t; t_i) \quad \forall i = 1, \ldots, l, \tag{9}$$

where $K(t;y)$ is the reproducing kernel of a Reproducing Kernel Hilbert Space (RKHS) H and $\{(t_i, y_i)_{i=1}^{l}\}$ is a data set, which has been obtained by sampling, in absence of noise, the target function $x_i(t)$ [15].

While Olshausen et al., in their overcomplete models, measure the reconstruction error with an L_2 criterion, the Support Vector Machine measures the true distance, in the H norm, between the target function and the approximating function. Depending on the value of the sparseness parameter, the number of coefficients a_{ij} that differ from zero will be smaller than r (the number of basis functions) (see (2)). The data points associated with the non-zero coefficients are called support vectors and it is these support vectors that comprise our sparse approximation.

3 Learning Objective

From the model presented in Sect. 2 we derive a simple and robust learning algorithm by maximizing the data likelihood over the basis functions. The learning objective is to find the most probable explanation for the input data. In other words, we wish to develop a generative model that encodes the probabilities of the input data. The algorithm must be able to:

- find a good matrix $\mathbf{M}$ for coding input data,
- infer the proper state of the coefficients $\mathbf{a}$ for each input signal.

In a special case of zero noise, in the data model presented in Sect. 2, and a complete representation (i.e. $\mathbf{M}$ is invertible), the problem leads to the well-known Independent Component Analysis algorithm [16]. ICA allows the learning non-orthogonal bases for data with non-Gaussian distribution. First we proposed to use ICA. Some data distributions, however, cannot be modeled adequately by either PCA or ICA.

The second and more substantial method is to place sparse prior constraints on the base probabilities of coefficient activation. This sparse coding constraint encourages a model to use relatively few basis functions to represent any specific input signal. If the data has certain statistical properties (it is "sparse"), this kind of coding leads to approximate redundancy reduction [17]. Sparse encoding within neural networks has previously been shown to create more biologically plausible receptive fields (Olshausen & Field).

3.1 Evolutionary Algorithm for proposed sparse coding

Some research has been done in applying genetic algorithms (GA) to the blind source separation (BSS) and ICA in which the Kullback-Leibler entropy is computed by analyzing the signals one by one and therefore needs exhausted computation [18].

In our work, an evolutionary algorithm (EA) is used to solve the problem of finding the best representation of a given signal in terms of basis functions and coefficients. EA searches for an optimum by changing iteratively population of temporary solutions encoded into chromosomes. Each chromosome represents the matrix of basis functions $\mathbf{M}$ and the matrix of coefficients $\mathbf{a}$. Fitness function minimized in our case is based on (8) and consists of two parts: 1) error of reconstructed signals and 2) sparse cost of the values of coefficients:

$$f = \sum_{i=1}^{m} \left(\sum_{t=1}^{n} \left| x_i(t) - \sum_{j=1}^{r} a_{ij} M_j(t) \right| + \sum_{j=1}^{r} S(a_{ij}) \right), \tag{10}$$

where $x_i(t)$ is the i-th input signal.

Genetic operators used in this algorithm are: crossover, mutation and macromutation. The crossover operator replaces, with *crossover probability*, each gene (number) in a chromosome with corresponding gene from another chromosome.

The mutation operator changes, with *mutation probability*, each number in a chromosome, by adding to it or subtracting from it a random value from *mutation range*.

The macromutation operator makes the same changes to the chromosome as mutation, but with higher values of *mutation probability* and *mutation range* parameters. This additional operator brings more diversty to the population and is used when there is no improvement of the best solution found so far. The main steps performed by EA to find the optimal set of basis functions are as follows:

```
Initialize the population of chromosomes
        with random values from range [-1, 1]
While the best fitness value found so far is not acceptable do:
   In each generation repeat:
      Calculate fitness for each individual
      Select best individuals
         according to the roulette selection rule
      Apply crossover and mutation
      Find best chromosome (lowest fitness value) in generation
      Save the best chromosome found so far
   If the difference between best chromosome found so far and
         the best chromosome found 20 generations ago equals 0,
      then decrease mutation range 2 times
   If mutation range is less than 0.00001,
      then apply macromutation and set mutation range
         to the initial value.
```

The algorithm can create representation of a given signal in terms of any number of basis functions.

When the matrix of basis functions or coefficients is not square, it is impossible to find an inversion of it, which would allow to compute representation of a new signal in terms of basis functions or coefficients already obtained. Such transformation can be done with the same evolutionary algorithm, but this time the chromosome consists only of coefficients or basis functions – whichever are being computed, and the fitness function is still (10).

The function in (10) is difficult to optimize due to large number of variables, which creates huge search space for the EA. Thus the problem can be divided into sub-problems in the following way.

Let us present (10) as:

$$f = \sum_{i=1}^{m} \left(\sum_{t=1}^{n} |E_j(t)| + \sum_{j=1}^{r} S(a_{ij}) \right), \tag{11}$$

where:

$$[E_j(t)] = \mathbf{E} = \mathbf{M}\mathbf{a} - \mathbf{x}. \tag{12}$$

The signal $\mathbf{x}$ can be divided along time, such that:

$$\mathbf{x} = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_k]^T. \tag{13}$$

Then, (12) can be written as:

$$[\mathbf{E}_1 \mathbf{E}_2 \dots \mathbf{E}_k]^T = [\mathbf{m}_1 \mathbf{m}_2 \dots \mathbf{m}_k]^T \mathbf{a} - [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_k]^T, \tag{14}$$

where:

$$[\mathbf{m}_1 \mathbf{m}_2 \dots \mathbf{m}_k] = \mathbf{M}, \tag{15}$$

$$[\mathbf{E}_1 \mathbf{E}_2 \dots \mathbf{E}_k] = \mathbf{E}, \tag{16}$$

and

$$\mathbf{E}_i = \mathbf{m}_i \mathbf{a} - \mathbf{x}_i, \ i = 1, 2, \dots, k. \tag{17}$$

Thus, having fixed the matrix of coefficients $\mathbf{a}$, all the parts of basis functions matrix $\mathbf{m}_i$'s can be computed independently. The pairs of $\mathbf{m}_i$ and $\mathbf{a}_i$ for each part of the input signals $\mathbf{x}_i$ can be computed using the evolutionary algorithm described above. Then the basis functions $\mathbf{M}_i$ corresponding to each coefficients matrices $\mathbf{a}_i$ for the whole input signals can be computed. The pair $\mathbf{M}_i$, $\mathbf{a}_i$, which gives the lowest value of the fitness function (10) is the obtained representation of the input signals. Therefore the algorithm is as follows:

```
Divide the signals into parts along time: x_1, x_2, ..., x_k
For each signal part x_i
   Compute with EA: m_i, a_i
For each a_i
   Compute with EA basis functions for the whole signal: M_i
Choose the pair a_i M_i that gives the smallest value
of fitness function (10).
```

4 Results and discussion

4.1 Data

In the experiments conducted at the Laboratory of Visual System, Department of Neurophysiology, Nencki Institute of Experimental Biology, Warsaw, Poland, a piezoelectric stimulator was attached to a vibrissa of a rat [19], [20]. An electrical impulse of 5 V amplitude and 1 ms duration was applied to the stimulator causing a deflection of the vibrissa. Evoked Potentials were then registered – each of them related to a single stimulus. Evoked potentials have been used from many years as a measurement of dynamic events occurring in nervous systems that accompany and are related to some defined sequences of behavior [12]. Based on same previous work, a hypothesis about a relation between two components of the registered evoked potentials and particular brain structures (i.e. supra- and infra-granular pyramidal cells) was stated. In order to verify the hypothesis, two additional types of stimuli were applied: 1) a cooling event applied to the surface of the cortex (allowing to temporarily "switch off" some structures of the brain), and 2) an aversive stimulus – electrical shock applied to the rat's ear (in order to cope with the phenomenon of habituation). Main goal of these experiments was to investigate those stimuli in the sense of their impact on the brain activity represented by evoked potentials.

A single, four-level electrode positioned in the cortex of a rat collected the data. The electrode registered brain activity in a form of evoked potentials on four depths simultaneously as described in [19]. The channels were defined as: channel 1 – 1.7 mm, channel 2 – 1.05 mm, channel 3 – 0.4 mm, channel 4 – surface. Each evoked potential (lasting $\approx$ 50 ms and separated by a 3 second period) was sampled with frequency of 2kHZ and thus is being described in the database by 100 values. The data sets contain 882 evoked potentials for each depth registered in the experiment, so the complete database consists of:

- four data sets for each channel
- 882 records in each data set
- each record described by 100 attributes (values in time)

Based on the description of the neuro-physiological experiments it is known which records (evoked potentials) correspond to a cooling event and *roughly* what the strength of this particular cooling event was.

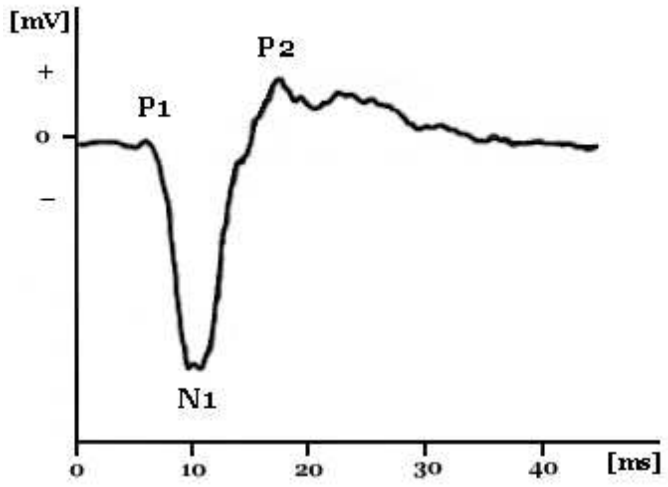

Fig. 1. Single (here averaged) evoked potential - one record in our database.

Sample (averaged) evoked potential from the database along with its division into three waves (two positive and one negative one) is presented in Fig. 1.

Because of the fact that the third channel's electrode (0.4 mm) was located in the closest position to the granular cells (laying in the middle between supra- and infra- granular, pyramidal cells – see [19], [20], [21]) and yielded the most representative perspective at the activity of the cortex, this level was quite often acknowledged as the most meaningful and interesting one and was given particular attention.

4.2 Analysis

The proposed model of sensory activity (the proposed model performance was analyzed via two methods. First, we applied the Independent Component Analysis (using the EEG/ICA Toolbox for Matlab by Scott Makeig, et al., [16]) to all the four channels. Second, the decomposition of EP by sparse coding and EA was analyzed.

Method 1: Averaged signals for all the four levels of the rat's cortex were used as an input to the algorithm (Fig. 2). Those averaged potentials were treated as four separate channels similarly to the traditional application of ICA to EEG data [16]. The full list of the ICA parameters is presented in Table 1.

Table 1. Independent Component Analysis parameters.

Parameter	Value
Channels (sources of activation)	4
Frames (points per one record)	100
Epochs	1
Sampling rate (in Hz)	2000
Limits (in milliseconds)	[0 50]

As a result, we received a 4 x 4 mixing matrix, which allowed us to decompose the input signals into four components. The components obtained by using this technique correspond very closely to the previous results derived from PCA (i.e. first two components create the N1 wave and their amplitudes change over time) [19], [20], [21]. Thus, we received a new representation of the signals in terms of:

- Time courses of activation of the Independent Components (Fig. 3)
- Independent Components (Fig. 4)

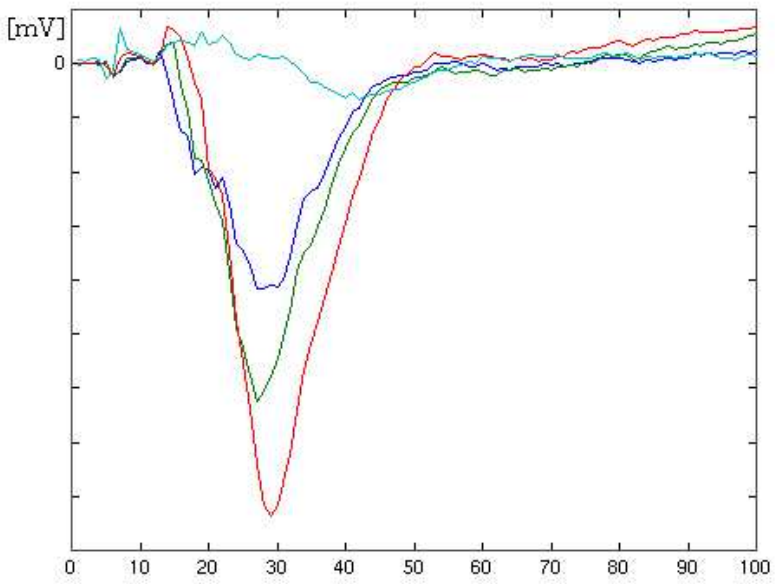

Fig. 2. Averaged signals for four separate channels.

This experiment has resulted in a new representation of the input signals in terms of their statistically independent components. These components, yielded

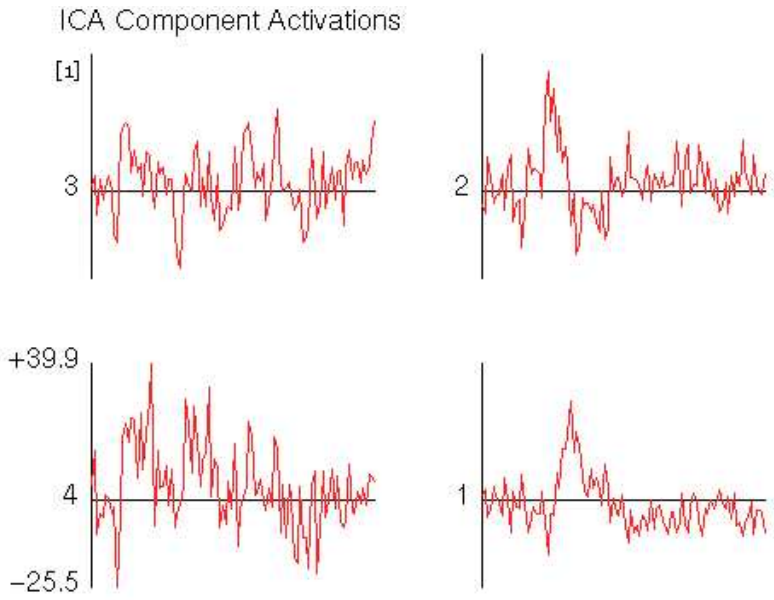

Fig. 3. Time courses of activation of all four components (ordered by latency).

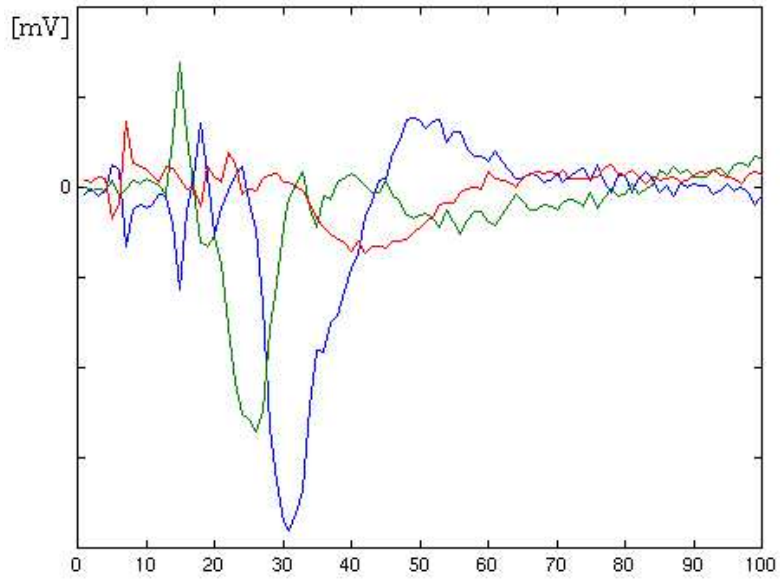

Fig. 4. Four independent components in the averaged 3^{rd} channel signal.

by an alternative technique (i.e. ICA), coincided with the ones discovered in previous work (i.e. PCA). This transformation (or coding) could be simply considered a data preprocessing methodology and, based on that, some further analysis could be performed. For instance, the values of the course of activation of the independent components (Fig. 3) could be used as new input data instead of the original signals and the analysis of the properties of such a new model might be of interest. On the other hand, generation of brand new data in the domain of the quantitative *description* of the ICA representation (e.g. minimum (maximum) value of each component, number of the minimum (maximum) component, etc.) would create another great possibility for investigation of the variability of the model's properties depending on the changing environment (i.e. cooling event).

Method 2: The experiment was performed in order to examine the effectiveness of the proposed algorithm in decomposing signals into components and compare the results with the ICA method. Normalized averaged signals for all levels of rat's cortex, treated as four separate channels, were used as an input to the algorithm (as in Method 1). The input signals are shown in Fig. 3. Initial values of the parameters of the algorithm are presented in Table 2.

Table 2. Initial values of the parameters of the proposed algorithm in this experiment.

Parameter	Value
Population size (number of chromosomes)	50
Crossover probability	0.5
Mutation probability	1 / no. of elements in $\mathbf{M}$
Mutation range	0.1
Macromutation mutation probability	0.5
Macromutation mutation range	0.01
Number of basis functions	10
σ parameter in (4)	0.3
β parameter in (4)	1.0

We have received 10 basis functions and 4 vectors of coefficients, which decompose the input signals. Three of the components of the averaged 3^{rd} channel signal, presented in Fig. 5 are very similar to those obtained by ICA. Slight differences in shapes and amplitudes are due to greater number of basis functions, which decompose the signals. One of those similar components seems to correspond to N1 wave.

Then, the representations of both, the averaged *normal* and *cooled* 3^{rd} channel signals were computed. The representation of the averaged *normal* and *cooled* 3^{rd} channel signals in the domain of the obtained basis functions (as values of coefficients) is presented in Fig. 6. Having more basis functions, more differences can be discovered between studied input signals. This may be very useful

in signal classification. Waves can be divided into more groups, depending on dissimilarities, not discovered in the domain of smaller number of basis functions.

The most important advantage of this method is the fact that number of basis functions is independent on the number of input signals.

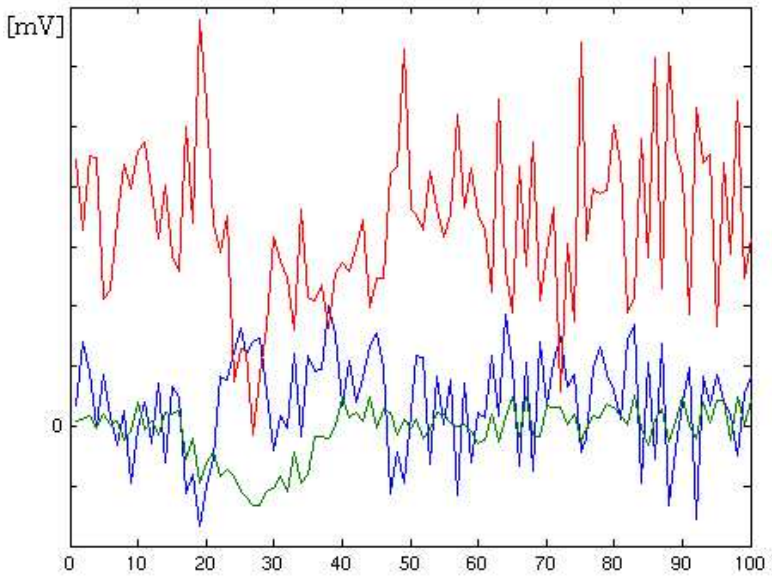

Fig. 5. Three arbitrarily selected components in the averaged 3^{rd} channel signal.

This evolutionary algorithm-based sparse coding of evoked potentials could be very useful in terms of some more detailed investigation of the input signals. For instance, it can be seen in Fig. 6 that the *normal* and *cooled* potentials have obviously different 6^{th}, 7^{th}, 8^{th}, and 10^{th} components. Such differences can be now thoroughly examined with some automatic data mining techniques like classification rules discovery or pattern matching in order to find some relations and dependencies between different kinds of evoked potentials.

5 Conclusions

On the basis of the experiments and the analysis described above we can conclude that ICA is quite a reasonable and effective tool in terms of evoked potentials' decomposition and transformation. The results, coherent with previous work in terms of the signal's main components, along with a high insensitivity of this method to noise and other types of distortion, encourage its further application for this type of problem. Using this clear and intuitive decomposition of signals, one may try to perform an investigation of the behavior of the components within

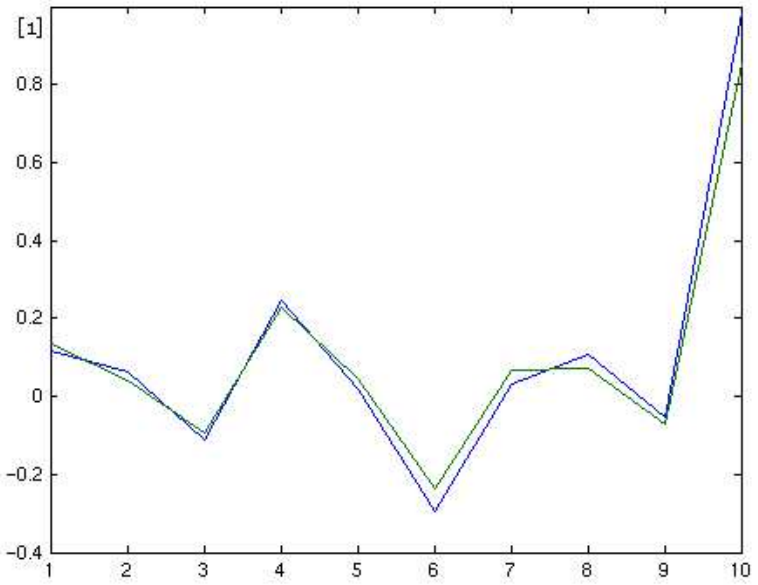

Fig. 6. Coefficients for the averaged *normal* and *cooled* 3^{rd} channel signals.

a changing environment of impulse stimuli, cooling and habituation that may be able to answer many questions about mechanisms ruling brain function.

Overcomplete representation, on the other hand, potentially allows a basis to better approximate the underlying statistical density of the data. It also creates the opportunity to discover more independent signal sources than the dimensionality of the input data, which opens new possibilities for data analysis. Along with the application of sparse coding constraints put on the learning algorithm, this is definitely a methodology worth of further exploration.

Both, the ICA- and the sparse coding-based modeling of evoked potentials seem to be a very reasonable and useful techniques for the data preprocessing. Based on those transformations, further investigation and analysis of the data via, for instance, classification or clustering algorithms is possible. Such algorithms, in turn, may allow researchers to discover quite new rules standing behind the mechanisms governing our brains' function.

References

1. Haykin, S.: Neural networks: a comprehensive foundation. 3rd edn. Prentice Hall, Upper Saddle River, NJ (1999)
2. Oja, E.: A simplified neural model as a principal components analyzer. J. of Mathematical Biology **15** (1982) 267–273
3. Amari, S.-I. and Cichocki, A.: Adaptive blind signal processing - neural network approaches. Proc. of the IEEE **86** (1998) 2026–2048
4. Barlow, H. B.: Possible principles underlying the transformations of sensory messages. In: Rosenblith, W. A. (ed.): Sensory Communication. MIT Press, Cambridge, MA (1961) 217–234
5. Raz, J., Dickerson, L., and Turetsky, B.: A Wavelet Packet Model of Evoked Potentials. Brain and Language **66** (1999) 61–88
6. Chen, S., Donoho, D. L., and Saunders, M. A.: Atomic decomposition by basis pursuit. Technical report, Dept. Stat., Stanford University (1996)
7. Lewicki, M. and Sejnowski, T.: Learning overcomplete representations. Neural Computation **12** (2000) 337–365
8. Mallat, S. G. and Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Trans. on Signal Processing **41(12)** (1993) 3397–3415
9. Olshausen, B. and Field, D. J.: Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision Research **37(23)** (1997) 3311–3325
10. Olshausen, B.: Sparse codes and spikes. In: Rao, R. P. N., Olshausen, B. A., Lewicki, M. S. (eds.): Probabilistic Models of Perception and Brain Function. MIT Press, Cambridge, MA (2001)
11. Milanova, M., Wachowiak, M., Rubin, S., and Elmaghraby, A.: A Perceptual Learning Model on Topological Representation. Proc. of the IEEE International Joint Conference on Neural Networks, Washington, DC, July 15-19 (2001) 406–411
12. Freeman, W. J.: Measurement of Cortical Evoked Potentials by Decomposition of their Wave Forms. J. of Cybernetics and Information Science, **2-4** (1979) 22–56
13. Lewicki, M. S. and Olshausen, B. A.: Probabilistic Framework for Adaptation and Comparison of Image Codes. J. Opt. Soc. of Am., **16** (1999) 1587–1600
14. Girosi, F.: An Equivalence Between Sparse Approximation and Support Vector Machines. Neural Computation **10** (1998) 1455–1480
15. Vapnik, V.: The nature of Statistical Learning Theory. Springer-Verlag, Berlin Heidelberg New York (1995)
16. Makeig, S., et al.: ICA Toolbox Tutorial.
Available: http://www.cnl.salk.edu/∼scott/tutorial/
17. Field, D. J.: What is the goal of sensory coding? Neural Computation **6** (1994) 559–601
18. Yoshioka, M. and Omatu, S.: Independent Component Analysis using time delayed sampling. Presented at the IEEE International Joint Conference on Neural Networks, Como, Italy, July 24-27 (2000)
19. Kublik, E. and Musial, P.: Badanie ukladow czuciowych metoda potencjalow wywolanych (*in Polish*). Kosmos **46** (1997) 327–336
20. Wrobel, A., Kublik, E., and Musial, P.: Gating of the sensory activity within barrel cortex of the awake rat. Exp. Brain Res. **123** (1998) 117–123
21. Kublik, E., Musial, P., and Wrobel, A.: Identification of Principal Components in Cortical Evoked Potentials by Brief Surface Cooling. Clinical Neurophysiology. **112** (2001) 1720–1725