The EEG Signal Classification in Compressed Sensing Space

Monica Fira
Institute of Computer Science
Romanian Academy
Iasi, Romania
e-mail: mfira@etti.tuiasi.ro

***Abstract*—In this paper, it is analyzed the possibility of the classification of the compressed sensed electroencephalographic (EEG) signals. Compressed sensing is a signal processing technique for efficiently acquiring and reconstructing a signal, by finding solutions to underdetermined linear systems. This is based on the principle that, through optimization, the sparsity of a signal can be exploited to recover it from far fewer samples than required by the Shannon-Nyquist sampling theorem. The signals classification is done directly in the compressed space and the EEG signals reconstruction is not necessary. For testing we used EEG signals from a brain computer interface system used for a spelling paradigm. For the classification task, two methods were used, both based on machine learning, namely, Deep learning and Gradient boosting learning.**

Keywords- EEG; Compressed sensing; BCI; classification; P300

I. INTRODUCTION

Compressed sensing (CS), during recent years, was in focus for various fields of science and technology as applied mathematics, computer science, electrical engineering, and signal processing. The novelty introduced by CS is that, in certain conditions, the traditional limits of sampling theory can be overcome. CS relies on the fundamental fact that various signals can be represented using only several nonzero coefficients in a suitable basis of vectors. This basis is named dictionary. Based on the known dictionary and using only very few measurements, such signals can be reconstructed using nonlinear optimization methods. Compressed sensing method is an example of usage in practice of recent mathematical results [1] – [4].

The literature from recent years comprises an impressive number of papers in the field of CS, including 1D and 2D medical signals. Among 1D signals, the most frequently analyzed in connection with CS applications are ECG and EEG since they are most used in the medical world. In the case of EEG signals, there is often a need of records for longer periods of time (i.e., during the night) or for a large number of channels. On the other hand, during the past few years, the human-computer interaction has been thoroughly investigated by researchers from the fields of neurology, psychology and information technology [1] – [8].

Over the past decades, the development of the technology of brain–computer interface (BCI) has provided a novel and promising communication channel for patients suffering from severe motor disabilities, but, being cognitively intact, they need an alternative method to interact with the environment.

As a non-muscular communication and control system, BCI has shown emerging possibilities for people with severe motor disabilities by allowing them to write sentences, move a cursor on the computer screen, play an electronic ping pong game, control an orthosis that provides hand grasp, or operate a brain actuated wheelchair. During the last two decades, BCI electroencephalographic (EEG) based systems have used a variety of electrophysiological signal components: visual evoked potentials, slow cortical potentials, P300 evoked potentials, mu and beta rhythms, and cortical neuronal action potentials [9] [10].

The P300 is a characteristic waveform in the human EEG, occurring as a response to rare task-relevant stimuli in a series of task-irrelevant stimuli. The classical oddball paradigm is usually used to evoke the P300: two categories of stimuli are presented to a subject in random order, one of the categories occurs only rarely and subjects are instructed to determine to which category a stimulus belongs [9] – [11].

The main goal of this paper is to test the following hypothesis: if some data can be classified in the original space, they can be acquired using the CS principle and then they can be classified with approximately similar results in the compressed space. In other words, the close neighbors remain close and far neighbors remain far in compressed space. In other words, the proportion of distances between neighbors is preserved.

The layout of the rest of the paper is as follows: In Section II, there are described the principles of compressed and the mathematical formalism of this method. In Section III is presented P300 spelling paradigm of brain computer interface. In Section IV is described the experimental paradigm used for this work, the subjects and the preprocessing of the data. In Section V, the boosting algorithm and deep learning used for classification are described. The experimental results and conclusions are presented in Section VI.

 ISBN: 978-1-61208-571-5

II. COMPRESSED SENSING

Compressed sensing is a rather new paradigm in signal processing that speculates the fact that the so-called sparse signals can be reconstructed from a sma.ll number of projections on a set of random signals. CS applications in biosignals acquisition, compression and processing have been intensely investigated in the last decade [1] – [7].

It is well-known that data acquisition is fundamentally governed by the familiar sampling theorem [8] that states that an f_0-bandlimited signal can be recovered from its samples if the sampling frequency is at least 2f0, i.e., twice the highest frequency of the signal spectrum. Thus, in a time window W, an f_0-bandlimited analog signal can be represented by $N=2f_0W$ samples equally spaced at $T=1/2f_0$, i.e., as a vector belonging to the space R^N. Such a signal can be alternatively described using any complete set of orthogonal functions in R^N. Let us observe that sampling is equivalent to taking projections (scalar products) on the elements of the *canonical* basis. In the general case, the signal can be reconstructed from its projections on N orthogonal (or only linear independent) elements in R^N the canonical basis being the most natural particular case and usually the most convenient. Indeed, the above considerations are rigorously valid in the tacit hypothesis of an infinite precision sampling.

On the other hand, there are many (classes of) signals that allow reconstruction based on fewer samples or projections than those required by the sampling theorem. The explanation is that in such cases the samples contain redundant information so that the signals can be compressed and can be reconstructed using projections and *known prior information*. Such a class is that of sparse signals, the *prior* information about them being the fact they admit a representation based on a small number of elements/atoms in R^N. A signal is called k-*sparse* if it is known that it can be represented using a number k of elements of R^N, the most interesting case being that when k<<N.

Formally, a discrete signal/vector $x \in R^N$ is said to be k-sparse if there exists a basis $\Psi = \{\Psi_i, i = 1,..., N\}$ in R^N such that most of the elements $\alpha = \{\alpha_i, i = 1,..., N\}$ of its representation in that basis, $x = \Psi\alpha$, are zero or, in a more relaxed hypothesis, *approximately zero* so that the signal can be represented *well enough* with the k's largest terms α_i from its expansion with respect to the above basis. In other words, x is said to be compressible, since it can be represented only by the nonzero/largest elements α_i. CS theory shows that a k-sparse signal, i.e., which is compressible in a base (or, more general, dictionary) Ψ can be recovered with very good quality from a number *m* of the order of magnitude

$$m = O(k \log(N / k))$$

of non-adaptive linear projections on a set of vectors Φ which are not coherent with the first, i.e., their elements cannot be used for a compressed representation of any $\Psi_i, i = 1,..., N$.

Thus, instead of measuring the N components of the signal in the canonic base, a number of m $(k < m << N)$ linear projections on the elements of the matrix Φ^{N*m} are acquired for obtaining the measurement signal (show in Figure 1)

$$y = \Phi x = \Phi\Psi\alpha = \Theta\alpha \tag{1}$$

where we have neglected the measurement noise. The vectors on which x is projected onto are arranged as the rows of a mxN projection matrix Φ, m < N, where N is the size of x and m is the number of measurements.

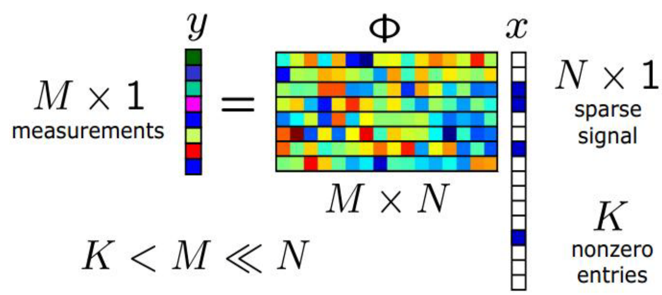


Figure 1. Matrix operations in Compressed Sensing

The system of equations (1) is obviously undetermined since $m < N$, the reconstruction of the initial signal can be made only based on the hypothesis that it is compressible. Under certain assumptions on Φ and Ψ, however, the original expansion vector α can be reconstructed as the unique solution to the optimization problem

$$\hat{\alpha} = \arg\min \|\alpha\|_{l_0} \quad \textit{subject to} \quad y = \Phi\Psi\alpha \tag{2}$$

the signal is then reconstructed with

$$\hat{x} = \Psi\hat{\alpha} \tag{3}$$

where l_0 is the pseudonorm equal to the number of nonzero elements of α, i.e., (2) amounts to finding the sparsest decomposition of the measurement vector y in the dictionary $\Phi\Psi$. Unfortunately, (2) is combinatorial and

 ISBN: 978-1-61208-571-5

unstable when considering noise or approximately sparse signals. Two directions have emerged to circumvent these problems: (i) pursuit and thresholding algorithms that seek a sub-optimal solution of (2) and (ii) the Basis Pursuit algorithm [1] that relaxes the l_0 minimization to l_1, solving the convex optimization problem (4) instead of the original one:

$$\hat{\alpha} = \arg_{\alpha} \min \|\alpha\|_{l1} \quad \textit{subject to} \quad y = \Phi\Psi\alpha \quad (4)$$

Many of the results obtained so far in CS refer to “genuine” sparse signals, i.e., to signals that can be represented using precisely k<<N atoms from a given dictionary. However, the results are formally valid for signals that are “approximately sparse”, i.e., k is the number of non-negligible elements. Moreover, signals can be sparse in overcomplete dictionaries Ψ, i.e., dictionaries with more atoms than their dimension; certain biomedical signals have been found to be sparse in such kind of overcomplete dictionaries. This is the reason why in the past few years, techniques inspired from the mathematic fundamentals of CS have also been applied in the field of biomedical signals, both at the level of processing methods for electroencephalographic (EEG) signals [2] – [4] but also in practical applications [5] including compression, transmission and reconstruction of ECG signals using smart-phones [6].

III. BRAIN COMPUTER INTERFACE - P300 SPELLER PARADIGM

P300 speller paradigm uses the P300 waves that are expressions of event related potential produced during decision making process. P300 has two subcomponents (as shown in Figure 2 a): the novelty P3 (also named P3a), and the classic P300 (renamed as P3b). P3a is a wave with positive amplitude and peak latency between 250 and 280 ms; the maximum values of the amplitude are recorded for the frontal/central electrodes. P3b has also positive amplitude with a peak around 300 ms; higher values are recorded usually on the parietal areas of the brain. Depending on the task, the latency of the peak could be between 250 and at least 500 ms.

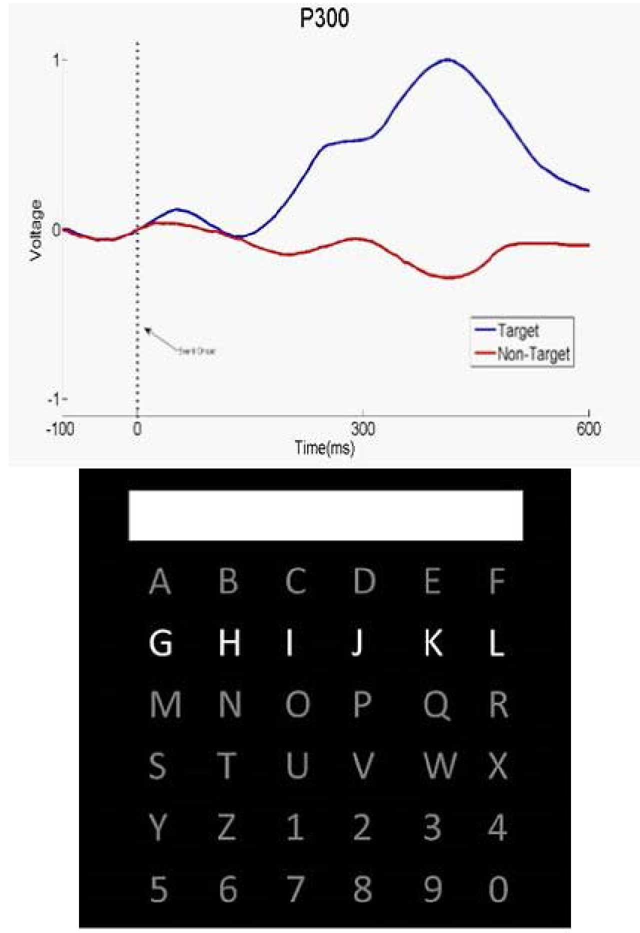


Figure 2. P300 wave and the classical P300 spelling paradigm described by Farwell-Donchin 1988

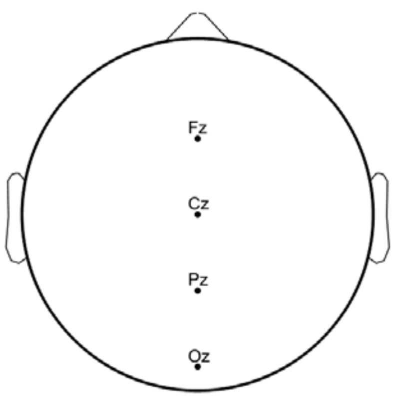
One of the first examples for BCI is the algorithm proposed by Farwell and Donchin [11] that relies on the unconscious decision making processes expressed via P300 in order to drive a computer. In their approach, a 6x6 matrix (see Figure 2.) of symbols is presented to the user and rows and columns of the matrix are flashed in random order. Subjects can select a symbol from the matrix, by counting the number of times it flashes. Each time the desired character flashes, a P300 is elicited and can be detected by an appropriate algorithm.

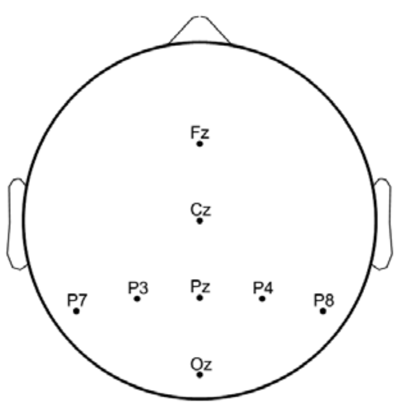
IV. EXPERIMENTAL PARADIGM, DATA ACQUISITION AND PREPROCESSING

In order to test the classification methods, we used EEG signals collected by Ulrich Hoffmann and collaborators in their laboratory and used by them in the papers [11][15]. The EEG data are available on the internet free at [14].

A setup similar to that described in [11] was used to record and to label the data. A 6×6 matrix containing the letters of the alphabet and the numbers 1-9 was presented to the subjects on a laptop screen. Rows and columns of the matrix were flashed randomly for 100ms with a 100ms pause between flashes. Flashes were block randomized, i.e., after 12 flashes each row and column was flashed exactly one time. Two datasets were recorded from each of the subjects on different days. In the first session subjects were asked to spell the french words ”lac,” ”nuage,” ”montagne, ”and ”soleil.” In the second session subjects had to spell the words ”fromage,” ”chocolat,” ”pain,” and ”vin” [12].

Data was recorded from channels FP1, FP2, AF3, AF4, F7, F3, FZ, F4, F8, FC1, FC5, FC6, FC2, T7, C3, CZ, C4, T8, CP1, CP5, CP6, CP2, P7, P3, PZ, P4, P8, PO3, PO4, O1, OZ, O2 with a Biosemi Active 2 system at 2048Hz. The data was then re-referenced to the average of channels O1, OZ, O2, lowpass filtered between 0 and 9 Hz with a 7th order Butterworth filter, and downsampled to 128 Hz. The channels used for re-referencing and channels T7, T8 were not used for further computations because they did not carry relevant information for the detection of P300s. A more detailed description for experimental paradigm, data acquisition, and preprocessing and artifact rejection is presented in [12].





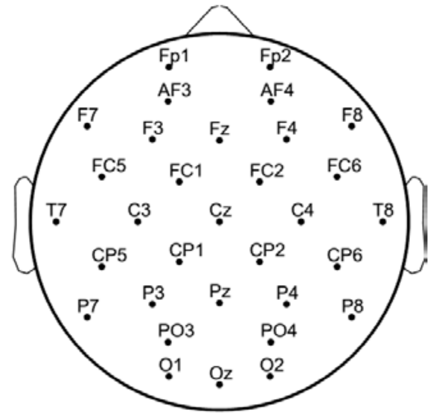


Figure 3. Electrodes configurations

In this paper, it was used a small dataset from this databased with EEG signals.

V. The Classification Methods

In this section the boosting algorithm and deep learning used for classification are described.

A. Gradient boosting

The Gradient boosting classifier from [12] was used. It should be noted that the used software was developed by the authors of that work in order to make a comparison of classification results compared to the original EEG data and the same EEG data but suppose they were purchased used compressed sensed principle.

Gradient boosting is a machine learning method, which builds one strong classifier from many weak classifiers [12].

In [12], Hoffmann and collaborators have described a simple, yet powerful method to detect the P300 from single EEG trials and use it to build a P300 based spelling device. Boosting was employed, to compute from training data a function that detects P300s from single EEG trials. In particular, gradient boosting was used to stepwise maximize the Bernoulli log-likelihood of a logistic regression model. Here ordinary least squares regression was used as weak learner. For a gradient boosting with OLS presented in detail see paper [12].

B. Deep learning

In this paper, we used deep learning in order to learn useful representation of features directly from data. Autoencoder neural networks are able to extract features from unlabeleld data. Autoencoders are used as instruments to train deep neural networks. The training mechanism for autoencoders is considered unsupervised because no labeled data are needed. Autoencoders are trained to replicate their inputs at their outputs by finding a set of weights minimize the corresponding cost function: the error between the inputs and their reconstruction at the outputs [13]. An autoencoder has two parts: an encoder and a decoder. Both, the encoder and decoder could have multiple layers, but, usually, they are designed with a single layer for each of them [13]. The

 ISBN: 978-1-61208-571-5

training algorithm for autoencoders is back-propagation based.

By cascading two or more autoencoders, a deepnetwork can be obtained.

VI. Experimental Results and Discussions

The experimental results of classification will be presented in both real acquisition space and in compressed space where the EEG signal will be collected by applying the compressed sensed concepts.

For testing there were used EEG signals collected by Hoffmann, namely, a reduced database available at [13]. This database contains EEG signals collected for 32 channels, grouped in 942 vectors to be classified EEG, lasting 1 sec each.

We have chosen four electrode placement configurations, which we tested for both the original signal and the compressed EEG signals (see Figure 3).

In order to test the classification in the compressed space, we chose two compression ratios, namely, compression of 5: 1 and respectively 10:1. Thus, using compressed sensed algorithm and a random matrix, we simulated that we acquire compressed sensed EEG signals.

For compression evaluation we used the compression rate (CR) defined as the ratio between the number of bits needed to represent the original and the compressed signal.

$$CR = \frac{b_{orig}}{b_{comp}}$$

where b_{orig} and b_{comp} represent the number of bits required for the original and compressed signals, respectively.

A. Gradient boosting

For testing using the gradient boosting method, the configuration parameters were kept the same as in [12]. Namely, the maximal number of iterations of the boosting algorithm Mmax was set to 200, the optimal M was determined in a 30×10 cross-validation loop, and ε was set to 0.05.

TABLE I. The Maximum Classification Rate for Original and Compressed Sensed EEG Signals For Gradient Boosting

Gradient boosting method – 23 channels (Fp1, AF3, F3, Fc1, Fc5, C3, CP1, CP5, P3, Pz, PO3, PO4, P4, CP6, CP2, C4, FC6, FC2, F4, AF4, Fp2, Fz, Cz)	
The classification Space	**Max Classification rate**
EEG originals	86%
CS with 10:1	80%
CS with 5:1	79%
Gradient boosting method – 8 channels (Fz, Cz, Pz, Oz, P7, P3, P4, P8)	
EEG originals	86%
CS with 10:1	80%
CS with 5:1	79%
Gradient boosting method – 4 channels (Fz, Cz, Pz, Oz)	
EEG originals	81%
CS with 10:1	75%
CS with 5:1	73%

Figure 4 shows the accuracy obtained during the cross-validation loop for configuration with 23 channels. As can be seen, the gradient boosting algorithm converges to an optimal solution. The difference between the rate of classification in the original and compressed sensed space is relatively small, only 6 percent.

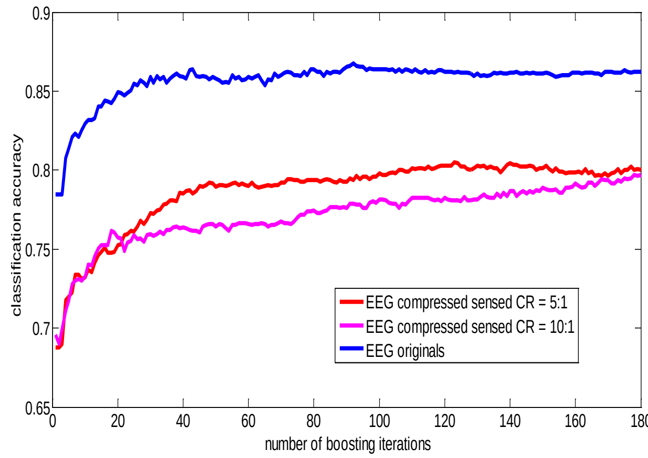


Figure 4. The percentage of classification performance for different values of M.

B. Deep learning

The relevant features for P300 waves are not directly identifiable on each segment of signal we have available. From this reason we selected as tool for classification the deep networks with auto-encoders which are able to extract relevant features from unlabeled data.

TABLE II. The Maximum Classification Rate for Original and Compressed Sensed EEG Signals for Depp Learning

Deep learning neural network			
The classification Space	**Max Classification Rate %**	**Size of signals to be classified**	**Optimum network config.**
Gradient boosting method – 23 channels (Fp1, AF3, F3, Fc1, Fc5, C3, CP1, CP5, P3, Pz, PO3, PO4, P4, CP6, CP2, C4, FC6, FC2, F4, AF4, Fp2, Fz, Cz)			
EEG originals	95%	2944	200-50
CS with 10:1	81%	294	50-10
CS with 5:1	78%	588	65-10
Gradient boosting method – 4 channels (Fz, Cz, Pz, Oz)			
EEG originals	80%	512	50-5

 ISBN: 978-1-61208-571-5

CS with 10:1	74%	52	20-5
CS with 5:1	72%	104	50-5

We used deep networks consisting in two auto-encoders followed by a soft-max layer for the classification of original signals and also for the compressed sensed ones. In case of not compressed signals, the first auto-encoder has 200 hidden elements and the second one has 40. For the compressed signals, we used a first auto-encoder with 40 hidden elements and a second one with only 10.

VII. CONCLUSIONS

In this paper, it was analyzed the possibility of EEG signals classification (from a spelling paradigm) into EEG signal containing P300 waveform and EEG signals without P300 wave. This classification is the essential element in a BCI system for spelling. Thus, starting from a method proposed by Hoffmann which is based on the gradient boosting classification, it was tested the possibility of the classification of the compressed sensed EEG signals. In other words it was analyzed the possibility of classifying compressed EEG signals, into compressed space which was named compressed sensed space. The utility of this classification is derived from the fact that using the mathematical foundations of CS, the EEG signals can be acquired directly in a compressed form (i.e. the number of samples in the EEG signal is lower than the one indicated by the sampling theorem based on the Nyquist frequency).

It was noticed that using the gradient boosting algorithm, the obtained classification rates have close values for the normal space and the compressed sensed space. Thus, if the original EEG signals classification rate is 86%, for the CS space the classification rate is only 6% lower.

We studied also the classification possibility by using neural network of deep learning type and the results in terms of classification rate are very similar to the gradient boosting method.

The obtained results with both tested methods confirm the hypothesis presented in introduction, according to which the close neighbors in initially space remain close also in the compressed sensed space. This allows the classification of the signals acquired directly in compressed way and it is useful in the applications where only the membership class is important for a signal and not its shape.

ACKNOWLEDGMENT

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-0832 ***"Medical signal processing methods based on compressed sensing; applications and their implementation."***

REFERENCES

[1] S. Chen, D. Donoho and M. Saunders, "*Atomic Decomposition by Basis Pursuit*", SIAM Review, 43 (2001), pp. 129-159

[2] L. F. Polania, R. E. Carrillo, Manuel Blanco-Velasco and Kenneth E. Barner, "*ECG compression via matrix completion*", EUSIPCO 2011

[3] L. F. Polania, R. E. Carrillo, Manuel Blanco-Velasco and Kenneth E. Barner, "*Compressed sensing based method for ECG compression*", 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011

[4] C. Wang, J. Liu and J. Sun, „*Compression algorithm for electrocardiograms based on sparse decomposition*", Front. Electr. Electron. Eng. China 2009, 4(1): 10–14

[5] M. Fira and L. Goras, "*Comparison of inter-and intra-subject variability of P300 spelling dictionary in EEG compressed sensing*", International Journal of Advanced Computer Science and Applications , Vol. 7, No. 10, 2016

[6] M. Fira, "*Compressed Sensing of Multi-Channel EEG Signals: quantitative and qualitative evaluation with Speller Paradigm"*, International Journal of Advanced Computer Science and Applications, Vol. 7, No. 6, 2016

[7] N. Cleju, M. Fira, C. Barabasa and L. Goras, „*Robust reconstruction of compressively sensed ECG patterns*", ISSCS 2011 (The 10-th International Symposium on Signals, Circuits and Systems), 30 June – 1 July 2011, Iasi, pp. 507-510, 2011

[8] C. E. Shannon, "*Communication in the presence of noise*", Proc. Institute of Radio Engineers, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[9] J. R.Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller and T. M. Vaughan, "*Brain-computer interfaces for communication and control*", Clin. Neurophysiol. 113 (6), 767-791, 2002

[10] M. A. Lebedev and M. A. Nicolelis, "*Brainmachine interfaces: past, present and future*", Trends in Neurosciences 29 (9), 536 – 546, 2006

[11] L.A Farwell and E Donchin, "*Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials*", Electroencephalography & Clinical Neurophysio,70(6):pp 510-523, 1988

[12] U. Hoffmann, G. Garcia, J.-M. Vesin, K. Diserens, T. Ebrahimi, "*A Boosting Approach to P300 Detection with Application to Brain-Computer Interfaces*", Proceedings of IEEE EMBS Conference on Neural Engineering, 2005

[13] https://www.mathworks.com/help/nnet/ref/trainautoencoder.html#buyr01b-1 (22 May 2017)

[14] http://mmspg.epfl.ch/cms/page-58322.html (22 May 2017)

[15] U. Hoffmann, J. M. Vesin, T. Ebrahimi and K. Diserens, "*An efficient P300-based brain-computer interface for disabled subjects*", J Neurosci Methods. 2008 Jan 15;167(1):115-25. Epub 2007 Mar 13.