Natural image reconstruction from brain waves: a novel visual BCI system with native feedback

Grigory Rashkov[1,2], Anatoly Bobe[1,2], Dmitry Fastovets[3], Maria Komarova[3]
[1] Neurobotics LLC, Moscow, Russian Federation
[2] Neuroassistive Technologies LLC, Moscow, Russian Federation
[3] Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russian Federation

Abstract

Here we hypothesize that observing the visual stimuli of different categories trigger distinct brain states that can be decoded from noninvasive EEG recordings. We introduce an effective closed-loop BCI system that reconstructs the observed or imagined stimuli images from the co-occurring brain wave parameters. The reconstructed images are presented to the subject as a visual feedback. The developed system is applicable to training BCI-naïve subjects because of the user-friendly and intuitive way the visual patterns are employed to modify the brain states.

1. Introduction and related work

Currently, the usage of EEG-based BCIs in assistive and rehabilitation devices mostly comes down to the following scenarios:

1) using synchronous BCIs (based on event-related potential registration, e.g. P300) for making discrete selections;
2) using asynchronous BCIs based on motor imagery potential or concentration/decocentration-driven mental states for issuing voluntary controlling commands.

Both scenarios have some advantages which are, unfortunately, overweighed with severe limitations that hinder implementations of BCI technology in real-world tasks. Thus, in synchronous BCI paradigms, a wide variety of stimuli, including visual categories, can be utilized to explore and measure the evoked responses of a particular subject [1]. However, the whole set of stimuli has to be successively presented to the subject each time to determine his intention, which makes such approach inconvenient for the applications requiring fast, real-time control of an external device. Motor-imagery or other asynchronous BCIs do not require any external stimuli presentation, which allows a subject to produce voluntary mental commands at his own wish. At the same time, the ability of different subjects to perform various mental tasks is variable and depends on their personal physiological parameters and experience [2]. A typical asynchronous BCI scenario requires an unexperienced subject to undergo a long-time training routine to master the control of at least 2 or 3 mental states. The practical instructions on how to perform abstract mental tasks are often unclear to novice BCI operators, which adds to overall complexity of the procedure. As a consequence, different studies show that classification rates are highly inconsistent even for similar paradigms [3–5].

A solution could be to join the advantages of the two scenarios by exploring the effect of continuous stimuli presentation on brain wave patterns. The decoded long-term evoked responses, if any, could then be treated as subject-specific mental states which could potentially be triggered by a particular stimuli imagination. One suggestion is to use natural movies of different objects as attention-capturing continuous stimuli. This kind of stimulation has been already reported for fMRI research [6]. Grigoryan et.al. have recently showed the positive effect of mobile visual stimuli in EEG-based BCIs as well [7].

Another essential part of most BCI protocols is maintaining a neurofeedback loop for the subject. Many articles have shown that the self-regulation effects achieved through this approach could facilitate learning of mental states [8–12]. Ivanitsky et.al. figured out an important role of BCI feedback even for mastering

complex cognitive tasks [13]. Ideally, a feedback should give a subject sufficient information about his progress, but not distract him from performing the mental task itself, i.e. being native in terms of the BCI paradigm. In recent works Horikawa et.al. [14, 15] propose a model for natural image reconstruction from fMRI brain signal data recorded from a subject while he observes the original images. There have been similar reports on similar EEG-based reconstructions [16], but the reliability of the studies has met serious controversy [17].

Considering the conditions outlined above, one can reasonably suggest that closed-loop asynchronous BCIs with adaptively modifiable set of mental states and a native type of feedback could outperform the other BCI approaches. In this article, we introduce a novel BCI paradigm that meets these requirements. Our protocol features the visual-based cognitive test for individual stimuli set selection as well as state-of-art deep learning based image reconstruction model for native feedback presentation.

2. Methods

For our research we set two major objectives:

- Exploring the continuous effect of visual stimuli content on rhythmical structure of subject's brain activity.
- Developing a model for mapping the EEG features extracted for a given category of observed stimuli back into the natural image space of the same category.

2.1. Subjects

The human protocol was approved by the local Institutional Ethical Committee (#5 of 18.05.2018). We recruited 17 healthy subjects with no history of neurology diseases, 11 males, 6 females, all right-handed. The subjects' age ranged from 18 to 33 years (mean age: 22). The subjects were informed about the experimental protocol and signed the informed consent form.

2.2. EEG recordings

The EEG recording equipment included a 128-channel EEG cap and NVX-136 amplifier developed by Medical Computer Systems Ltd. (Moscow, Zelenograd). The conductive gel (Unimax gel) was injected into each electrode. Each subject was seated in a comfortable position at a distance approximately 0.6m from an LCD computer monitor. The EEG signals were recorded at a sampling rate of 500 Hz with NeoRec software (developed by Medical Computer Systems Ltd.). The recorded signals were filtered using a band pass filter with a bandwidth of 1-35Hz.

2.3. Visual stimuli

In this study, we used video clips of different objects as stimuli. We assumed that observing the videos rather than the static images would keep the subjects motivated during the experiment, making an extra mental task unnecessary. We selected following stimuli object categories, which could potentially affect the subjects' mental state by inducing relaxation, concentration on particular items, or imposing stress (the examples are shown on Figure 1):

- A - abstract geometric shapes or fractals (visual illusions);
- W - natural waterfalls;
- HF - human faces with different emotions;
- GM - Goldberg mechanisms (mechanical devices with a large number of elements triggering each other);
- E - extreme sports (first-person videos of high speed motion activities, some ending with accidents).

The experiment consisted of two sessions with a short break of 5-10 minutes between them. During each session, a subject was asked to watch a 21-minute video sequence comprised of 117 randomly mixed

video clips. The duration of a single clip was between 6 and 10 seconds, and the “black screen” transitions of 1-3 seconds were inserted between the consecutive clips.

Figure 1. Examples of video frames from each of the stimuli categories

The video sequence contained 25 clips of each category except A, where only 17 fragments were present due to low variability of the category and the fatigue effect caused by such videos. There were no identical clips within a category, as well as no montage transitions within a single clip to avoid the occurrence of parasite ERPs. The onsets of the video clips were precisely synchronized with EEG signal using MCS VGASens photo sensor.

2.4. Feature extraction and classification

In order to simulate the real-world procedure of subject training, we used the first session for training the model and the second session for performance validation. The data was epoched into time segments corresponding to each single clip observation, and each time segment was split into 3-second time windows with 2/3 overlap.

Independent component analysis (ICA) matrix was calculated on the whole training session data. All of the signal processing stages including muscular and ocular artifact rejection and feature extraction were performed within ICA space. We used fast Fourier transform to extract the spectral features for each component remaining after artifact removal. As dense electrode recordings produce a way excessive feature dimensionality, we considered a scoring procedure for feature space compression. The average power spectrum values were obtained for each *k*-th data sample of each *n*-th component using a sliding frequency window *w* of 3 Hz:

$$\overline{PSD_{n,k,w}} = \frac{1}{f_{high} - f_{low}} \sum_{f_{low}}^{f_{high}} {S_{n,k,f}}^2$$

A set of simple threshold classifiers was created separately for each of the PSD values. The features were scored according to these classifiers performance on the training set. As each classifier gave an estimation of predictive power related to a particular frequency band of a particular ICA component, the components with the best overall feature scores were selected and the most informative feature bands were specified for each of the selected components.

The dimensionality of the feature vectors was then further reduced to a fixed value of 20 using principal component analysis (PCA) transform. The feature classification was performed using linear discriminant analysis (LDA).

3. Model evaluation

Table 2 presents the rates for pairwise category classification task. For each pair of classes, we balanced the number of features by randomly dropping some samples from the class that contained more data. The significance level for each binary classifier was estimated with a one-tailed binomial test. As we used overlapping windows for evaluation, we used the corresponding number of actually independent trials for setting the unbiased classifier confidence rate. The actual and recalculated number of samples and the corresponding confidence rates for non-random hypothesis acceptance are shown in Table 1.

Table 1. Confidence rates for different binary classifiers

	Number of oversampled features (for classifier)	Number of independent features (for binomial test)	Non-random classification rate (p=0.95)
For “A-W”, “A-HF”, “A-GM”, A-E”	320	68	0.61
For Others	614	130	0.57

The classification accuracy scores below the confidence rates are marked as gray cells in Table 2. The overall results were evaluated using two metrics: the probability for the given binary classification to be non-random for a subject (P_{nr}) and average classification accuracy among the subjects with non-random classification rate (P_{rec}). 73.5% classifications proved to be at least non-random, with average score of 77.3% among them.

We performed two-tailed Wilcoxon rank-sum tests for each pair of categories with significance level of $\alpha = 0.01$ to evaluate cross-subject stability of the results. The p-values and the q-values obtained after the Bonferroni-Holm correction are shown in Table 2. As a result of results analysis we decided to discard the “abstract shape” class for further experiments although it showed some promising results for some subjects.

Table 2. Classification results

Subj	A-W	A-HF	A-GM	A-E	W-HF	W-GM	W-E	HF-GM	HF-E	GM-E
1					75%	87%	84%	82%	79%	70%
2	62%		81%			81%		78%		79%
3	75%		85%	75%	75%	76%	79%	91%	75%	83%
4	73%	70%	74%	65%		74%	70%	86%	85%	77%
5			85%	73%		83%	66%	89%	87%	74%
6	75%		78%	74%	67%		80%	84%	76%	
7			89%	77%		83%	79%	91%	76%	83%
8	79%		87%		70%	76%	68%	86%	67%	77%
9	71%	71%	82%	70%	68%			67%	76%	71%
10					68%	70%		80%		85%
11		86%			66%	87%	77%	89%	80%	76%
12			71%		65%			70%	67%	
13	77%	69%	81%	67%		80%	74%	84%	84%	69%
14	71%	78%	84%		61%	79%	79%	90%	90%	78%
15	78%	68%	78%	79%	80%	84%	84%	86%	81%	80%
16		69%		77%		87%	76%	94%	90%	80%
17	86%	75%	82%	73%	70%	78%	71%	77%	78%	
P_{nr}, %	58.8	47.1	76.5	58.8	64.7	82.4	76.5	100.0	88.2	82.4
P_{rec}, %	74.1	71.7	80.6	72.6	69.5	80.4	76.7	83.3	79.4	77.4
p-val	0.0051	0.0116	0.0015	0.0051	0.0034	0.0010	0.0015	0.0003	0.0007	0.0010
q-val	0.0153	0.0153	0.009	0.0153	0.0136	0.0080	0.0090	0.0030	0.0063	0.0080

Considering the significance of the reported results, we have to express some concerns about the “Goldberg mechanisms” category. Most of the video clips referring to this category involved small

moving particles which provoked the subjects to visually follow them. The corresponding EEG signal was contaminated by eye movement artifacts which, despite our artifact removal precautions, could potentially contribute some parasite discriminative features to the classifier. Thus, authors of this work abstain from claiming the absolute purity of achieved results for this category as long as that does not lead to the loss of generality of the described research.

Overall, the model evaluation results show that some visual stimuli categories seem to be generally useful for BCI purposes among most of the subjects, and some other can be individually picked up or rejected depending on the results of the proposed cognitive test. Within each category, an exemplar-level analysis can be carried out to select the most effective stimuli for a particular subject.

4. Native neurofeedback model

The general idea of the proposed visual feedback is to present the BCI classifier predictions in form of natural images, which should be as close to the actually observed (or imagined) visual stimuli as possible. This would boost the subject's imagination efforts by giving him an illusion of his thoughts being visualized and at the same time would not distract him with any side-triggers. A basic approach would be simply to show some image samples from the stimuli dataset according to the category predicted by the classifier. A strong drawback in this solution is that weak or uncertain predictions would be either strictly mapped to one of the categories or cause excessive image switching, both ways much confusing to the subject. We hypothesized that a proper visual feedback should satisfy the following criteria:

- dynamically represent the brain wave decoding process
- retain at least the general content of the reconstructed object so that a subject can always recognize the decoded category;
- avoid quick or sudden shifting between the category types even if the signal becomes noisy or inconsistent;
- represent classifier uncertainty in adequate form of "uncertain" image

In order to keep in terms with these requirements we developed a deep-learning based visualization model. The general scheme of the model is shown on Figure 2. The 20-dimensional EEG feature vector obtained after dimension reduction stage (see "*Feature extraction and classification*" section above) is mapped into the latent space of a pre-trained image autoencoder, which is capable of reconstructing natural images of several pre-learnt categories. An image decoding model is independent of any neurophysiological data and can be trained beforehand considering just a set of stimuli images. A feature mapping model is trained separately as it requires both EEG feature bank and a trained image decoder at use. In the following sections we explain the model training procedure in more detail.

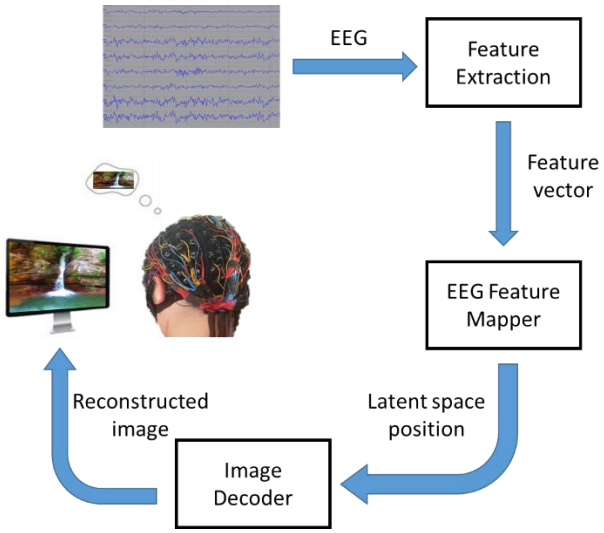
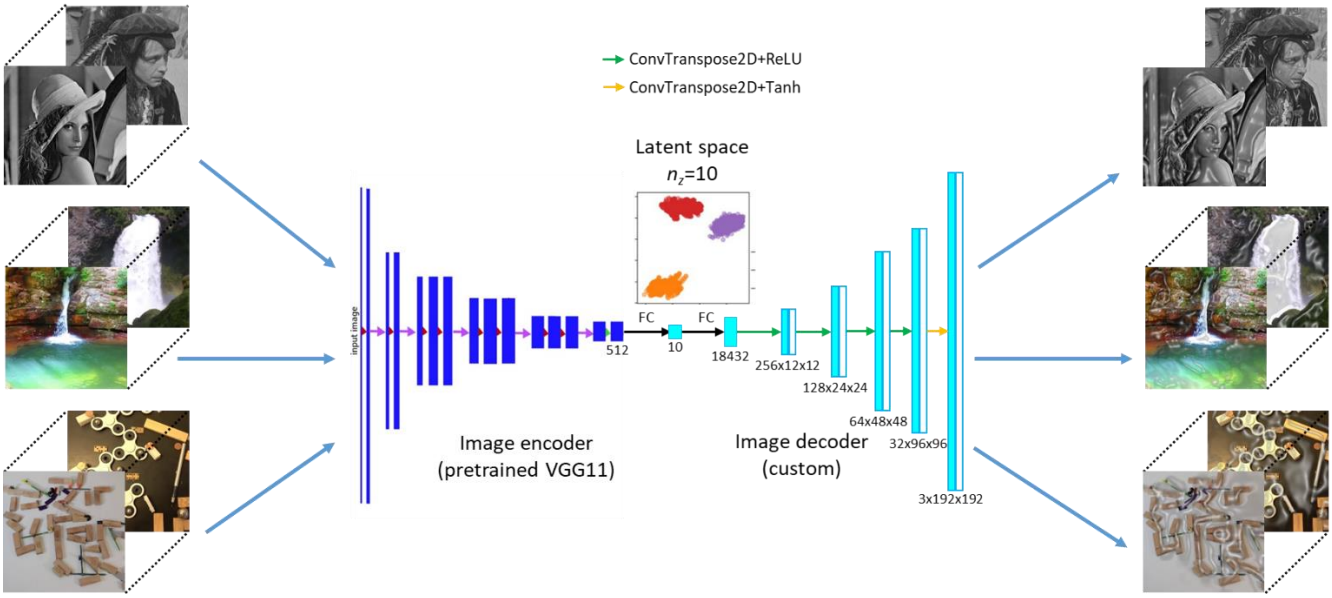


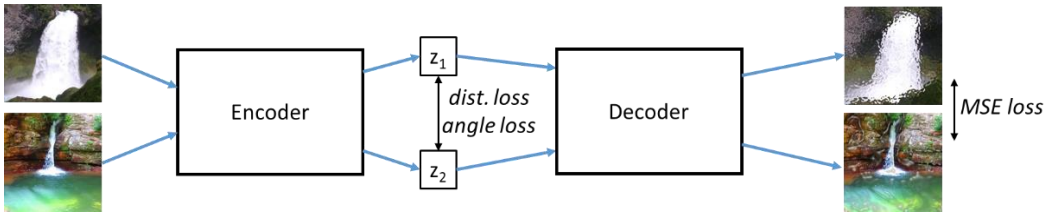
Figure 2. General scheme of a neurofeedback model

4.1. Image Decoder

An image decoder (ID) is a part of image-to-image convolutional autoencoder model. The encoder part is based on a pre-trained VGG-11 model [18]. The decoder part is composed of a fully-connected input layer for dimension enhancement, followed by 5 deconvolution blocks, each one containing a deconvolutional layer followed by rectifier linear unit (ReLU) activation. The final deconvolutional block contains hyperbolical tangent activation layer. A decoder produces color images of 192x192x3 dimensionality (see Figure 3a).



a



b

Figure 3. Image decoder. a) Model structure; b) Training routine

Apart from image reconstruction, we suggest our decoder model to have a specific distribution of its latent space. We handled this problem by introducing a training procedure shown on Figure 3b. The encoder and the decoder were in a "siamese" fashion. A siamese network can be seen as two identical subnetworks with shared weights [19]. Each network processes its own input sample and the weights are updated according to a contrastive loss function, so that a model learns to judge whether the inputs belong to the same class or not. For our model the goal is to translate the visual similarity measure between a pair of input images I_1, I_2 into the mutual distance between a pair of corresponding vector representations z_1, z_2 in n-dimensional latent space. Moreover, the vector clusters should be compact for each of the image categories and also should be uniformly spread across the latent space to prevent the occurrence of large "blank gaps" which would affect the quality of reconstructed images I_{1r}, I_{2r}. Considering all the specified features, we proposed a loss function as a weighted sum of three components: distance loss, angle loss and pixel loss. In this work after some parameter tuning we set the weights as w_d=1, w_a=4, w_p=2.

$$L = w_d L_d + w_a L_a + w_p L_p$$

Distance loss function was used to control the mutual distance between latent space representations and was calculated as following:

$$L_d = td + (1-t) \cdot \mathrm{S}^2(m-d)$$

$$t = |l - e^{\eta d}|$$

$$d = \frac{1}{n}\sqrt{\sum (z_1 - z_2)^2}$$

where l=1 if the images belong to the same category, otherwise l=0. Therefore, target coefficient t is close to zero for similar images of the same category;

S is sigmoid function:

$$\mathrm{S}(x) = \frac{1}{1+e^{-x}}$$

η is a distance weighting parameter (in this research we used $\eta = 10^{-3}$);

m is margin that separates clusters in latent space. Here we used m=1.

In angle loss function cosine similarity metric was used to maintain the uniformness of cluster positions across latent space:

$$L_a = (1-t) \cdot \frac{z_1 \cdot z_2}{\|z_1\| \cdot \|z_2\|}$$

This loss function prevents the category clusters from forming a linear distribution in the latent space, making them form a kind of polygon instead. The mutual distances between the cluster centroids in such distribution is more or less similar, and no a priori preferences are given to any class.

Pixel loss is a common loss function for generative model loss that controls the quality of image reconstruction. For a pair of images, it is a sum of mean square errors for both image reconstructions:

$$d = \frac{1}{N}\sqrt{\sum (I_1 - I_{1r})^2} + \frac{1}{N}\sqrt{\sum (I_2 - I_{2r})^2}$$

Where N is number of pixels in the images. Another useful feature of a pixel loss is that it makes similar images from the same video clip (e.g. faces of the same person) to have close locations in latent space, therefore the stimuli data tends to form compact subclusters within a common category cluster (see Appendix 1). For our feedback model it means that similar EEG features will be decoded into similar stimuli exemplars (e.g. a particular face) rather than switch chaotically between different exemplars of the category.

The image decoder was trained on a dataset comprised of the image frames taken from the training session video for subject-specific preselected categories. Image pairs were randomly selected to create equal number of same class pairs and different class pairs. Some visualizations for ID model performance can be found in Appendix 1.

4.2. EEG feature mapper

The aim of an EEG feature mapping network (FM) is to translate the data from EEG feature domain (f) to the image decoder latent space domain (f'). Ideally, an image observed by a subject and the EEG recorded at the time of this observation would finally be transformed into the same latent space vector, so that a decoder would produce a proper visualization of what the subject had just seen or imagined:

$$\mathrm{FM}(f) = f' \approx z$$

Another problem is to cope with noisy or inconsistent data: while EEG signal properties in real-time recording scenario can significantly vary due to the undetected artifacts or subject getting distracted, the

feedback system still should be kept from chaotic image switching as that would put excessive stress on the subject.

The fact that we are working with continuous data gives ground for utilizing the recurrent network models for solving the task. In this research we used long-short term memory (LSTM) cells as recurrent units [20]. We also incorporated the attention mechanism, which makes the model emphasize the key data features and ensures better stability against outliers [21]. A scheme of our EEG feature mapper is shown on Figure 4a and its training routine is presented on Figure 4b.

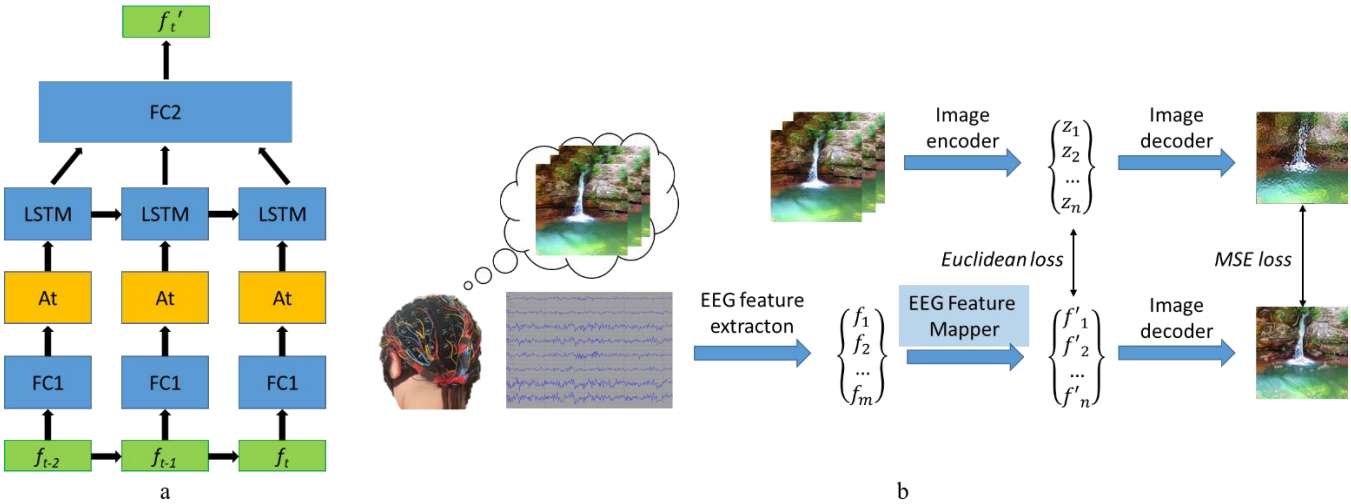


Figure 4. EEG feature mapper. a) Model structure; b) Training routine

A loss function for feature mapper model minimizes the mean square error between EEG and image feature representations both in latent space and in image reconstruction space after decoding:

$$d = \frac{1}{n}\sqrt{\sum (z - f')^2} + \frac{1}{N}\sqrt{\sum (I_r - I'_r)^2}$$

$$I'_r = \mathrm{ID}(f')$$

The feature mapper was trained on a dataset comprised of the image frames from the training session video and corresponding 3-second EEG signal windows (centered on the moment of frame onset). EEG feature vectors were extracted using the same method as described in Feature extraction section of this work. Some visualizations for FM network performance can be found in Appendix 1.

4.3. Real-time evaluation and results

The proposed neurofeedback model was implemented on Python (with deep learning models implemented using pyTorch library) and run on a machine with an Intel i7 processor, NVIDIA GeForce 1050Ti GPU and 8 Gb RAM. The EEG data was collected in real time via lab streaming layer (LSL) data flow from the amplifier. The processing speed was nearly 3 frames per second, which included incoming EEG data acquisition, filtering, feature extraction and image reconstruction. Before passing to the real-time experiments we extensively tested our neurofeedback model in emulation mode using the data sessions recorded for different subjects. We trained the feedback model on three categories which contributed maximal mutual classification rates for each particular subject. Unfortunately, the objective criteria for feedback visualization quality is yet to be developed, and we had to rely on subjective judgements considering the achieved results. Typically, around 90% the reconstructed imaged were recognizable in terms of category affiliation. The correct correspondence between the classifier predictions and the reconstructed image object category was always established in cases when classifier prediction was more

or less stable for a period of 2-3 seconds. This delay time can be regulated by FM network hidden state parameter and can be decreased, which comes at cost of losing the image stability.

The examples of reconstructed images for 4 object categories are shown on Figure 5. More examples can be found in Appendix 1.

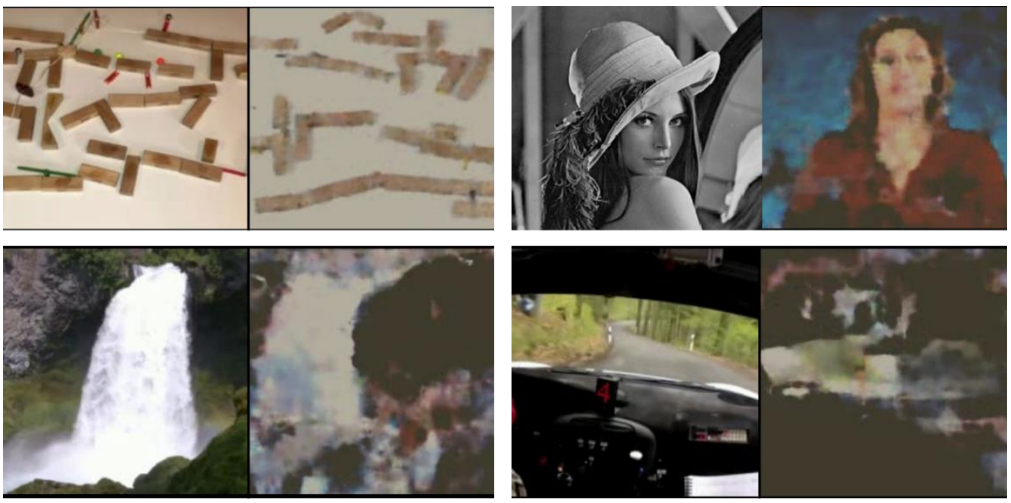


Figure 5. Original images from video stimuli and reconstructed images obtained after processing the co-occurring EEG signal (an original face image replaced by an image sample due to publication policy)

For testing our neurofeedback system in real time we requested one of the subjects to watch the experimental video from the test session once again (Figure 6). The model performance was found to be the same as at the emulation stage. We also requested the subject to watch the presented feedback images instead of the actual video clips and try to switch between the categories at his own will. The subject managed to master imaginary switching between two of three offered categories during 10 minutes of training.

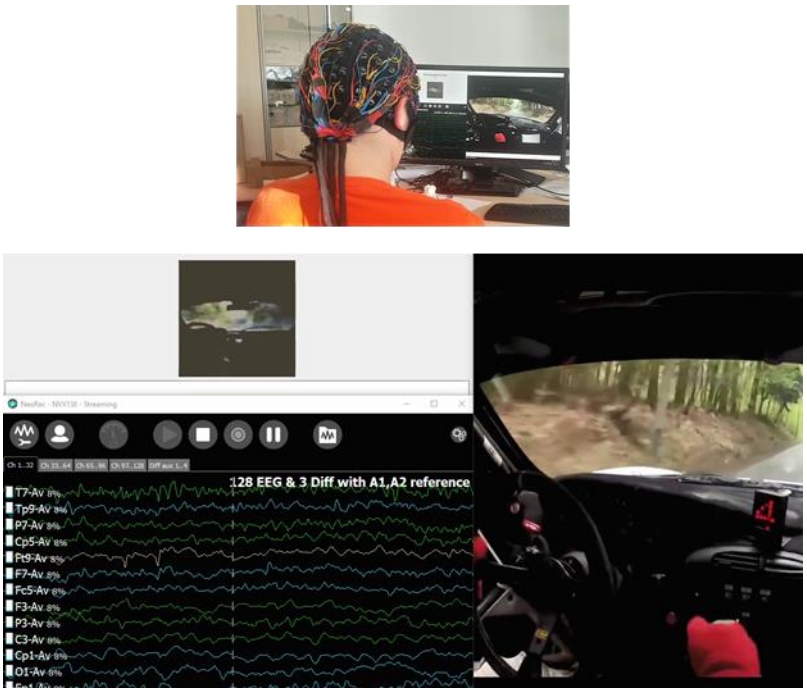


Figure 6. Real-time experimental run. On upper left corner of the screen a reconstructed image of its original (192x192) size can be seen. For this illustration we positioned the EEG recording software window on the same display as the neurofeedback output. During evaluation we hid this window to minimize the distracting factord for the subject.

5. Discussion

In this research we explored the effect of continuous visual stimuli presentation on the subject's brain waves registered in dense noninvasive EEG. The obtained results show that the electrical brain activity can be modulated by presenting subjects with the visual stimuli in the form of video clips. Different classes of objects present in the video clips caused different effect on the brain electric potentials, making it possible to distinguish the discrete scalp EEG patterns which corresponded to each of the object classes and were stable in time. We showed that it was possible to vary the stimuli within each category without affecting the inter-category separability. This makes the proposed method suitable for non-ERP based synchronous BCI implementations.

Our stimulation protocol can be considered as a cognitive test that aims to extract subject-specific stable EEG patterns. Each person has specific reactions to different kinds of visual stimuli, thus, for an effective BCI paradigm a preliminary step of individual stimuli set selection from some excessive basic set could be of a great value. Another benefit of this approach is that no additional cognitive task for attention or memory is required, and the subject can remain completely passive throughout the session. This makes it possible to implement this protocol for the patients with cognitive disorders.

Basing on the results achieved through the developed experimental protocol, we proposed a novel closed-loop BCI system, which is capable of real-time image reconstruction from the subject's EEG features. We developed a deep learning model which consists of two separately trained deep learning networks, one of which is used for decoding of different categories of images, and the second one transforms the EEG features into the image decoder spatial domain. We demonstrated that the proposed technique can potentially be used for training BCI-naïve subjects by replacing the original stimuli with the subject's mind-driven image reconstruction model. We suggest that using native feedback could produce a strong self-regulating effect and help a BCI operator to master the imagery commands more effectively.

Further extensive research is required to explore the efficiency of the BCI system in real-world applications. In our future work we will focus on following aspects:

- exploring more visual stimuli categories;
- improvement of reconstructed images quality by incorporating discriminator into image decoder network;
- developing a criteria for assessing the quality of images reconstructed from EEG features
- setting up comparative experiments to evaluate the role of adaptive stimuli selection and visual feedback presentation in BCI operator training speed and quality.

6. Funding

This work was accomplished through financial support of National Technological Initiative Fund (NTI), grant №5/17 from 12.05.2017.

7. Acknowledgments

The authors would like to thank Dr. Georgiy Ivanitsky and Prof. Alexei Ossadtchi for the useful discussions related to the details of experimental protocol and EEG data processing.

References

[1] Mensen A, Marshall W, Tononi G. (2017). EEG Differentiation Analysis and Stimulus Set Meaningfulness. Front Psychol; 8:1748. doi: 10.3389/fpsyg.2017.01748.

[2] Heremans, E., Vercruysse, S., Spildooren, J., Feys, P., Helsen, W. F., and Nieuwboer, A. (2013). Evaluation of motor imagery ability in neurological patients: a review. Mov. Sport Sci. / Sci. Mot 38. 31–38.

[3] Myrden, A., and Chau, T. (2015). Effects of user mental state on EEG-BCI performance. Front. Hum. Neurosci. 9:308. doi: 10.3389/fnhum.2015.00308

[4] Bashivan, P., Rish, I., and Heisig, S. Mental State Recognition via Wearable EEG. Proceedings of 5th NIPS workshop on Machine Learning and Interpretation in Neuroimaging (2015).

[5] Podobnik, B. (2017), NeuroPlace: Categorizing urban places according to mental states. PLoS One 12. doi: 10.1371/journal.pone.0183890

[6] Naselaris T., Kay K., Nishimoto S, Gallant J. (2011). Encoding and decoding in fMRI. Neuroimage. 15;56(2):400-10. doi: 10.1016/j.neuroimage.2010.07.073.

[7] Grigoryan R.K., Krysanova E.U., Kirjanov D.A., Kaplan A.Ya. (2018). Visual Stimuli for P300-Based Brain-Computer Interfaces: Color, Shape, and Mobility. ISSN 0096-3925, Moscow University Biological Sciences Bulletin, Vol. 73, No. 2, pp. 92–96.

[8] Gevensleben, H., Holl, B., Albrecht, B., Schlamp, D., Kratz, O., Studer P. et al. (2010). Neurofeedback training in children with ADHD: 6-month follow-up of a randomised controlled trial. Eur Child Adolesc Psychiatry 19, 715-724. doi: 10.1007/s00787-010-0109-5

[9] Sherlin, L., Lubar, J. F., Arns, M., and Sokhadze, T. M.: A Position Paper on Neurofeedback for the Treatment of ADHD J. Neurother. (2010) 14, p. 66–78. doi: 0.1080/10874201003773880

[10] Masterpasqua, F., and Healey, K. N. (2003). Neurofeedback in psychological practice. Prof. Psychol. Res. Pract. 34. 652–656. doi: 10.1037/0735-7028.34.6.652.

[11] Zapala, D., Wierzgala, P., Francuz P., and Augustynowicz, P. Effect of selected types of neurofeedback trainings on SMR-BCI control skills. in Conference: NEURONUS 2015 IBRO & IRUN Neuroscience.

[12] Marzbani, H., Marateb H. R., and Mansourian, M. (2016). Neurofeedback: A Comprehensive Review on System Design, Methodology and Clinical Applications. Basic Clin Neurosci. 7 p. 143–158. doi: 10.15412/J.BCN.03070208

[13] Atanov, M., Ivanitsky, G., and Ivanitsky, A. (2016). Cognitive brain-computer interface and the prospects of its practical application. Hum. Psychol. 42. p. 5-11

[14] Horikawa T., and Kamitani, Y. (2017). Generic decoding of seen and imagined objects using hierarchical visual features Nat. Commun. 8. doi: 10.1038/ncomms15037

[15] Horikawa, T. and Kamitani, Y. (2017). Hierarchical Neural Representation of Dreamed Objects Revealed by Brain Decoding with Deep Neural Network Features Front. Comput. Neurosci. 11:4, p. 1–26. doi: 10.3389/fncom.2017.00004

[16] Spampinato, C. Palazzo, S., Kavasidis, I., Giordano, D., Shah, M., Souly, N. (2016). Deep Learning Human Mind for Automated Visual Classification.

[17] Li, R., Johansen, J.S., Ahmed, H., Ilyevsky, T.V., Wilbur, R.B., Bharadwaj, H.M., Siskind, J.M. (2018). Training on the test set? An analysis of Spampinato et al.

[18] Simonyan, K., Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image

Recognition.

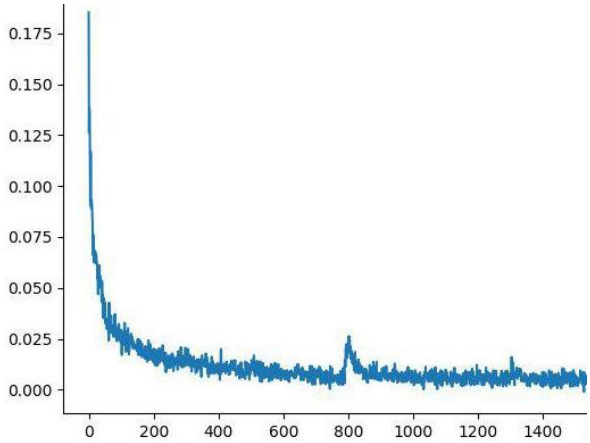
[19] Bromley, J., et al. (1994). Signature verification using a "siamese" time delay neural network. Advances in neural information processing systems.

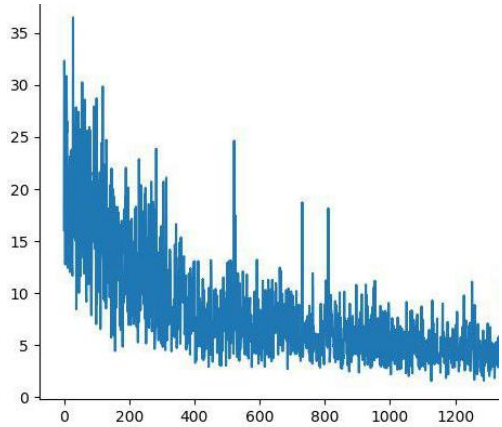
[20] Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8):1735–1780. doi: 10.1162/neco.1997.9.8.1735

[21] Kyunghyun, C., Courville, A., and Bengio, Y. (2015). Describing Multimedia Content using Attention-based Encoder–Decoder Networks. arXiv:1507.01053

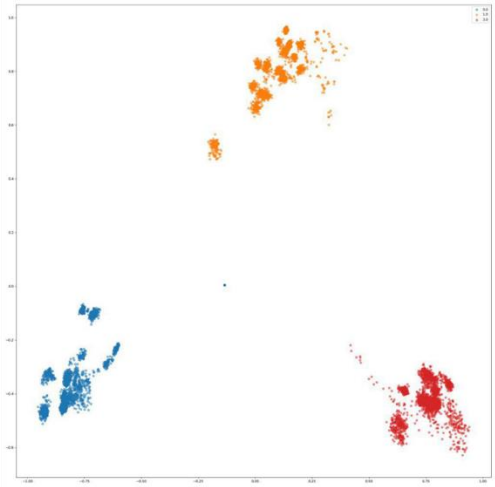
Appendix 1. Visualization of image reconstruction model performance*

*Here we present the results for three image categories. We do not present results for face image reconstruction due to the preprint publication policy restrictions.

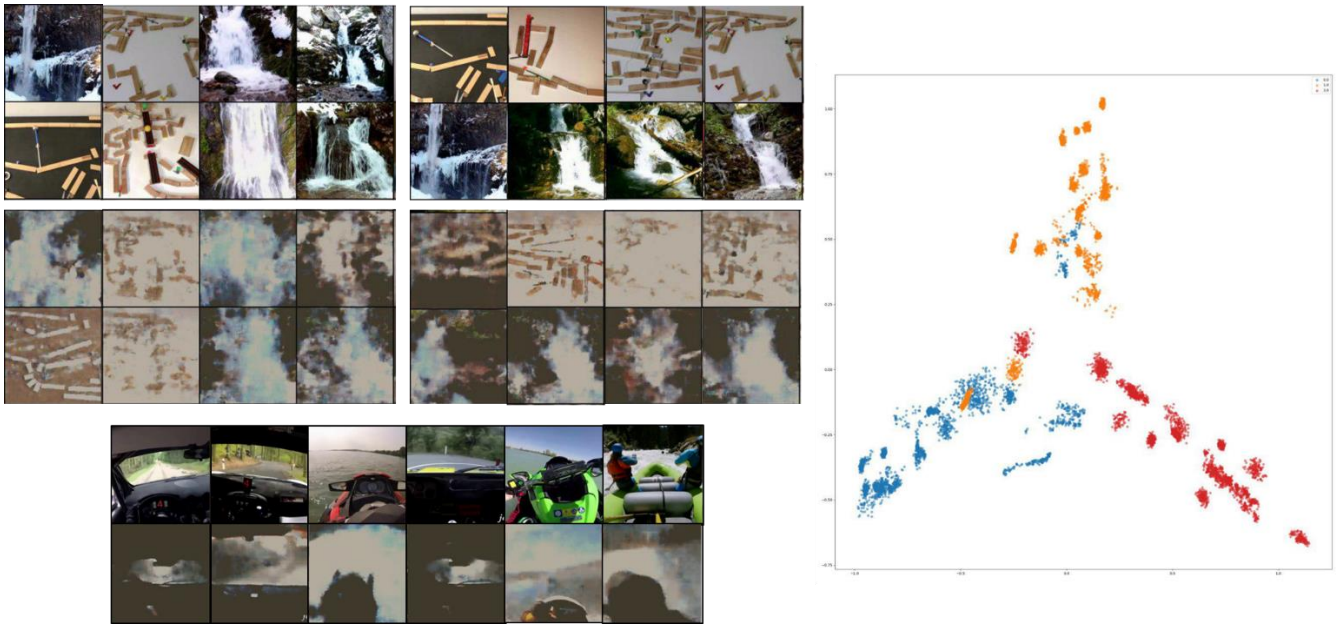




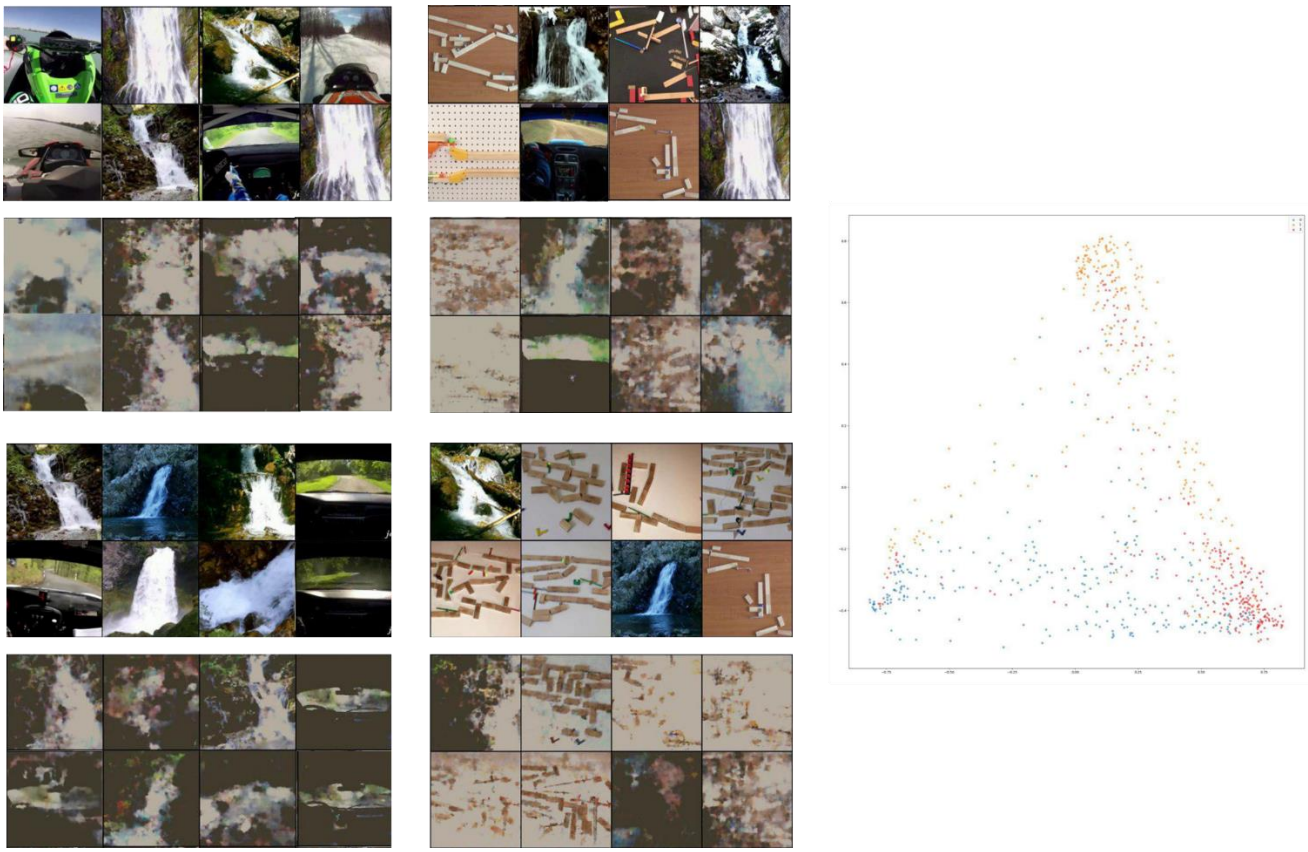
Learning curve of ID (left) and FM (right) models



ID performance examples on train data and cluster distributions in the latent space (compared to the corresponding original images from the stimuli video data set).



ID performance examples on test data and cluster distributions in the latent space. Note the inner-category subclusters formed by groups of similar images from single clips.



ID performance examples on EEG feature data mapped with FM and the distribution of the mapped data clusters in the latent space.