

National Research University Higher School of Economics

*as a manuscript*

Ossadtchi Alexei

**Methods for analysis, decoding and interpretation of  
brain's electrical activity in diagnostic,  
neurorehabilitation and neurotherapy applications**

DISSERTATION SUMMARY  
for the purpose of obtaining academic degree  
Doctor of Science in Computer Science

Moscow - 2022

The Dissertation for the purpose of obtaining academic degree Doctor of Science in Computer Science was prepared at the National Research University Higher School of Economics.

# Table of contents

<b>1</b>	<b>Main results</b>	<b>5</b>
<b>2</b>	<b>Introduction</b>	<b>8</b>
<b>3</b>	<b>New methods for solving the EEG and MEG inverse problem</b>	<b>14</b>
3.1	Adaptive beamformer with modified covariance matrix . . . . .	14
3.1.1	Introduction . . . . .	14
3.1.2	Data Model and Problem Statement . . . . .	15
3.1.3	New method . . . . .	18
3.1.4	Calculation and application of the projection operator . . . . .	18
3.1.5	Results . . . . .	21
3.2	Detection of functional networks with low phase delay from non-invasive measurements of brain activity . . . . .	23
3.2.1	Introduction . . . . .	23
3.2.2	Generating equation of cross-spectral matrix of sensor signals . . . . .	25
3.2.3	New method . . . . .	26
3.2.4	Results . . . . .	28
<b>4</b>	<b>New processing methods applied to the diagnosis of epilepsy</b>	<b>34</b>
4.1	Efficient biomimetic method for detecting interictal discharges (spikes) in multichannel recordings of brain electrical activity. . . . .	34
4.2	Traveling wave model for analysis of local dynamics of interictal discharges propagation and its application to determination of the epileptogenic zone . . . . .	40
4.2.1	Data Model . . . . .	40
4.2.2	Basic waves . . . . .	41
4.2.3	Optimal combination of traveling waves . . . . .	42
4.2.4	Main results . . . . .	42
4.2.5	Results on model data . . . . .	43
4.2.6	Results in these patients . . . . .	45
<b>5</b>	<b>New approaches to real-time processing and decoding of brain states for neural interfaces and neurofeedback systems</b>	<b>47</b>
5.1	Estimation of rhythmic activity parameters with minimal delay . . . . .	47
5.1.1	Mathematical model . . . . .	48
5.1.2	Existing methods . . . . .	49
5.1.3	Description of the developed family of methods . . . . .	49
5.1.4	Method comparison . . . . .	50
5.2	Decoding brain activity using interpretable neural networks . . . . .	53
5.2.1	Introduction . . . . .	53
5.2.2	Signal Observation Model . . . . .	53
5.2.3	Compact Convolutional Network Architecture . . . . .	55
5.2.4	Weights interpretation method . . . . .	56

5.2.5	Results . . . . .	57
5.2.6	Decoding finger kinematics from ECoG . . . . .	58
5.2.7	Speech decoding from invasive data . . . . .	59
5.2.8	Synchronous decoding . . . . .	62
5.2.9	Asynchronous decoding . . . . .	62
5.3	Network architecture . . . . .	63
5.4	Results . . . . .	63
<b>6</b>	<b>Conclusion</b>	<b>66</b>
	<b>References</b>	<b>69</b>

**Knowledge domain** Methods for analysis of data in natural sciences

**Aims and objectives** Improving the spatiotemporal resolution of brain function mapping through the use of physiologically determined models of neuronal activity in order to create new tools for research, diagnostics, neurorehabilitation and interaction with the brain.

## 1 Main results

- New methods for solving the inverse problem of EEG and MEG
  - An operation of projection in the  $R^N \times R^N$  product-space of vectors of magneto- and electroencephalographic (MEEG) data is proposed to ensure the stability of the beamformer method to neuronal sources with highly synchronized activations. In the realistic simulation mode, the properties of the method are investigated, a comparative analysis is carried out, and application to real data is demonstrated [Kuznetsova et al. \(2021\)](#); [Greenblatt et al. \(2005c\)](#)
  - A generalization of the projection operation in the  $R^N \times R^N$  product-space of a vector of magneto- and electroencephalographic (MEEG) data is proposed, in order to effectively reduce the volume conduction effect and detect functional networks with zero or small phase delay. In the realistic simulation mode, the properties of the method are investigated, a comparative analysis is carried out, and application to real data is demonstrated [Ossadtchi et al. \(2018\)](#); [Kuznetsova et al. \(2021\)](#); [Greenblatt et al. \(2012\)](#)
  - An iterative Bayesian method is proposed for increasing the spatial resolution of MEG inverse modelling by using interindividual similarities and differences in the spatial characteristics of neuronal activity. Within the framework of the approach based on optimization of the likelihood function of the second kind, a correlation matrix of a priori distribution of sources with non-strictly similar spatial characteristics between subjects is formed as a linear superposition of matrices that reflect the properties of similarity and difference of the individual cortical areas activity. Detailed modeling and a comparative analysis with other methods was carried out, and application to MEG data was demonstrated [Kozunov and Ossadtchi \(2015\)](#)
- New methods for processing ECoG, EEG, MEG as applied to the diagnosis of epilepsy
  - A computationally efficient biomimetic convolutional method for detecting interictal events (spikes) in multichannel recordings of brain electrical activity is proposed. This approach simplifies the doctor’s interaction with the algorithm and allows one to “code” the requirements for the morphology of the desired interictal spike in the form of a logical predicate. The resistance of the method to high-amplitude artifacts, which often accompany real recordings of interictal activity in patients with epilepsy, has been demonstrated [Kleeva et al. \(2022\)](#); [Ossadtchi et al. \(2004a, 2005\)](#).

- A method for assessing the cortical-wave dynamics of interictal discharges based on MEG data is proposed. The properties of the method were studied and its application to the interictal data of 9 patients was demonstrated. It is shown that the presence of interictal wave dynamics makes it possible to non-invasively judge the epileptogenicity of a cortical region [Kuznetsova and Ossadtchi \(2022\)](#). In combination with methods for analyzing the distal spread of interictal activity [Ossadtchi et al. \(2005\)](#), an arsenal of tools for non-invasive analysis of the electrophysiological activity of the brain is formed based on the analysis of the parameters of the distal and proximal spread of pathological cortical activity.
- A method for identifying epileptogenic networks based on invasive ictal (during a seizure) ECoG data is proposed. The approach is based on the estimation of pairwise functional connectivity of sensor signals and grouping into common spatial cliques of pairs of electrodes with similar profiles of changes in pairwise phase connectivity. The proposed approach for each of the analyzed seizures made it possible to identify dynamic networks and identify the primary network, the activation of which initiated the seizure, and the nodes (subset of electrodes) of which determined the epileptogenic zone [Ossadtchi et al. \(2010\)](#).
- New approaches to real-time processing and decoding of brain states for neural interfaces and neurofeedback systems
  - A compact architecture of a convolutional neural network for decoding multichannel electrophysiological signals and a theoretically justified method for interpreting its parameters to determine the geometric and functional properties of neuronal sources, pivotal for the downstream task [Petrosyan et al. \(2021a\)](#), are proposed.
  - Comparative analysis of achievable performance and interpretation accuracy was carried out and applications in [Petrosyan et al. \(2021b\)](#) speech and [Petrosyan et al. \(2021a\)](#) motor neurointerfaces [Lebedev and Ossadtchi \(2018\)](#) were demonstrated.
  - A family of xCFIR methods for assessing the instantaneous phase and envelope of rhythmic brain activity in real time has been developed, and a comparative analysis with SOTA [Smetanin et al. \(2020a\)](#) has been carried out.
  - Software has been developed for conducting experiments in the neurofeedback paradigm. The software has its own language for describing the configuration of the signal processing path, the artifact detuning block, the parameters of spatial and frequency filtering, and the features of presenting the feedback signal [Smetanin et al. \(2018a\)](#).
  - The application of this software to the problem of studying the effect of neurofeedback signal delay on the efficacy of neurofeedback training was demonstrated and it was found that reducing the delay leads to an increase in the learning rate and a longer retention of the training effects [Belinskaia et al. \(2020a\)](#); [Ossadtchi et al. \(2017a\)](#)

The author’s personal contributions include the theoretical formulation of new methods, the direct initial development and software implementation of the above approaches and algorithms, as well as full research management of theoretical and modeling analysis of performance characteristics and application to processing real data of electrophysiological activity in pathology and in norm and preparation of publications. More than 17 scientific articles and one book chapter have been published on the topic of this thesis in leading international journals. In all but one of the articles the applicant is the first or last author, and in one article (Greenblatt, R. E., Ossadtchi, A., & Pflieger, M. E. (2005)) the applicant has performed all analytical calculations and research of source localization estimation bias. All articles included in this dissertation research were published after the candidate defended his Ph.D. degree at the University of Southern California, Los Angeles, USA in 2003. During the last 5 years 5 Ph.D. theses have been written under the guidance of the applicant, 3 of them successfully defended in 2021, it is expected the defense of 2 more works by the end of 2022.

### **First-tier publications**

1. Kuznetsova, A., Nurislamova, Y., and Ossadtchi, A. (2021). Modified covariance beamformer for solving meg inverse problem in the environment with correlated sources. *Neuroimage*, 228:117677
2. Ossadtchi, A., Altukhov, D., and Jerbi, K. (2018). Phase shift invariant imaging of coherent sources (PSIICOS) from MEG data. *NeuroImage*, 183:950–971
3. Kozunov, V. V. and Ossadtchi, A. (2015). Gala: group analysis leads to accuracy, a novel approach for solving the inverse problem in exploratory analysis of group meg recordings. *Frontiers in Neuroscience*, 9:107
4. Greenblatt, R. E., Pflieger, M., and Ossadtchi, A. (2012). Connectivity measures applied to human brain electrophysiological data. *Journal of neuroscience methods*, 207(1):1–16
5. Greenblatt, R., Ossadtchi, A., and Pflieger, M. (2005a). Local linear estimators for the bioelectromagnetic inverse problem. *IEEE Trans Signal Proc*
6. Kleeva, D., Soghoian, G., Komoltsev, I., Sinkin, M., and Ossadtchi, A. (2022). Fast parametric curve matching (fpcm) for automatic spike detection. *Journal of Neural Engineering*, 19(3):036003
7. Ossadtchi, A., Baillet, S., Mosher, J., Thyerlei, D., Sutherling, W., and Leahy, R. (2004b). Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering. *Clinical Neurophysiology*, 115(3):508–522
8. Ossadtchi, A., Mosher, J. C., Sutherling, W. W., Greenblatt, R. E., and Leahy, R. M. (2005). Hidden markov modelling of spike propagation from interictal MEG data. *Physics in Medicine and Biology*, 50(14):3447–3469
9. Ossadtchi, A., Greenblatt, R., Towle, V., Kohrman, M., and Kamada, K. (2010). Inferring spatiotemporal network patterns from intracranial eeg data. *Clinical Neurophysiology*, 121(6):823–835
10. Petrosyan, A., Sinkin, M., Lebedev, M., and Ossadtchi, A. (2021a). Decoding and interpreting cortical signals with a compact convolutional neural network. *Journal of Neural Engineering*, 18(2):026019
11. Smetanin, N., Belinskaya, A., Lebedev, M., and Ossadtchi, A. (2020b). Digital

filters for low-latency quantification of brain rhythms in real time. *Journal of Neural Engineering*, 17(4):046022

12. Belinskaia, A., Smetanin, N., Lebedev, M., and Ossadtchi, A. (2020b). Short-delay neurofeedback facilitates training of the parietal alpha rhythm. *Journal of Neural Engineering*, 17(6):066012

13. Ossadtchi, A., Shamaeva, T., Okorokova, E., Moiseeva, V., and Lebedev, M. A. (2017b). Neurofeedback learning modifies the incidence rate of alpha spindles, but not their duration and amplitude. *Scientific reports*, 7(1):1–12

14. Smetanin, N., Volkova, K., Zabodaev, S., Lebedev, M. A., and Ossadtchi, A. (2018b). Nflab—a versatile software for neurofeedback and brain-computer interface research. *Frontiers in neuroinformatics*, 12:100

### Second-tier publications

15. Kuznetsova, A. and Ossadtchi, A. (2022). Анализ локальной динамики распространения межприступных разрядов с помощью модели бегущих волн. *Журнал высшей нервной деятельности им. И.П. Павлова*, 1(3):370–386

16. Petrosyan, A., Lebedev, M., and Ossadtchi, A. (2020b). Linear systems theoretic approach to interpretation of spatial and temporal weights in compact cnns: Monte-carlo study. In *Biologically Inspired Cognitive Architectures Meeting*, pages 365–370. Springer

17. Petrosyan, A., Lebedev, M., and Ossadtchi, A. (2020a). Decoding neural signals with a compact and interpretable convolutional neural network. In *International Conference on Neuroinformatics*, pages 420–428. Springer

## 2 Introduction

Scientists of various specialties have turned their efforts toward the object that is commonly believed to make us sentient beings, to provide our cognitive and analytical abilities, to shape our emotions, to enable us to create and to compactly accumulate knowledge about our world and our place in it. Such an assumption has lived on since Galen, who first formulated the idea that it is the brain that is the source of our thoughts. It would be wrong to underestimate the role of other organs in the processes of perception of information from the external world, formation of emotions and thoughts, but according to modern concepts, it is still the brain that scientists associate with our conscious behavior and analytical abilities, and also believe that its dysfunction is the cause for a number of neurological disorders. Whether this is true or not, and if so, to what extent, humanity has yet to find out, and for this we need both research tools and technologies to maintain and restore brain function, capable of prolonging the period of active intellectual life and thereby improving the quantity and quality of knowledge passed on from generation to generation.

At present, there is a wide range of tools for studying brain function: from the analysis of subjective questionnaires filled out by subjects and behavioral experiments to approaches that are based on technologies for objectively measuring brain’s activity. The latter can also be classified according to the type of activity being recorded. For example, positron emission tomography (PET) technology is based on recording the

intensity of metabolic processes and the principle that an increased concentration of a metabolite (glucose) characterizes the recently active part of the cortex. Functional magnetic resonance imaging (fMRI) monitors the variations in the concentration of oxygenated hemoglobin in different regions of the cortex, as well as in subcortical nuclei - neurons require energy to function, which is released through chemical reactions that require oxygen to flow. The disadvantage of these two methods is their low temporal resolution. Metabolic processes and changes in blood flow only accompany the fast-paced information exchange processes and replenish energy to the neuronal populations most involved in the immediate past.

Apparently, the main signaling mechanism ensuring information and computational processes in the brain is the generation by a neuron of an electric impulse called action potential (AP) discovered by Julius Bernstein and its transmission through synapses to the input of other neurons [Schuetze \(1983\)](#). Conglomerates of neurons interconnected by forward and backward connections form natural distributed neural networks capable of implementing almost any mathematical function. Information in such a network is encoded by the number of neuron PDs per unit time, and, in accordance with a number of theories, information exchange processes imply synchronization of neuron and neuron population activity. Generation of APs in nervous tissue occurs on a millisecond time scale. Thus, registration of electrical activity allows not only to study the neurophysiological basis of fast cognitive, motor and sensory processes and diagnose a number of pathologies, but also to create real-time systems for restoration or replacement of lost functions by solving the problem of decoding electrical activity of the brain and forming commands to external assistive devices, such as bionic limb prosthesis, exoskeleton or speech synthesis device. Such neural interface technologies have historically emerged as a separate field of applied neuroscience, which in the pursuit of funding has become overheated with promises of universal systems for reading and decoding brain activity with applications extending from obvious systems for motor function rehabilitation to futuristic mind reading devices, pumping information directly into the brain, direct communication between the brains of several people, etc. Nevertheless, it is the direction of functional neuroimaging that remains primary, much more knowledge-intensive, and has well-defined clinical applications.

Neuroimaging of fast processes implies registration brain's electrical activity and its processing using special algorithms in order to obtain dynamic (time-varying) maps of cortical or subcortical activity. The reading of the electrical activity of the brain can be performed invasively and non-invasively. Invasive technologies include methods of electrocorticography (ECG), stereo electroencephalography (stereo EEG), and intracortical electrodes - specialized matrices of needle electrodes penetrating into the cortex and capable of recording APs of individual neurons. Noninvasive registration of electrical processes occurring in the brain is performed using electro- and magnetoencephalography (EEG and MEG). Both methods involve the use of an array of sensors located on the surface of the scalp (EEG) or in close proximity (MEG) and measure, respectively, the electrical or magnetic field fluctuations generated by neuronal sources.

Not only neuronal sources, but also signals from muscle activity, eye movements, and cardiographic artifacts contribute to the electromagnetic field variability, both in invasive and noninvasive cases. Accordingly, to correctly interpret the measurements obtained, it is necessary to solve the problem of detuning from the above artifactual

signals.

When using invasive recordings, when reading electrodes form a direct contact with the nerve tissue, the problem of correlating the recorded activity with a certain area of the brain is usually solved in a trivial way and the corresponding area simply coincides with the position of the electrode.

In the case of noninvasive measurements, in order to construct maps of cortical activity it is necessary to solve an inverse EEG or MEG problem, which, like most problems of this class, is incorrectly posed. One class of approaches to solving such a problem is to apply regularization techniques, the point of which is to add a priori information about the properties of the activity being reconstructed on the cortical surface. Another family of methods, called the Local estimators method or the adaptive beamformer method (AFL), circumvents the incorrectness of the global problem by solving a set of tasks for estimating the activity of each individual cortical region. Further superposition of the estimates obtained for different cortical areas makes it possible to obtain a picture distributed over the entire cortex. This modification of the approach based on local solutions is called the scanning local estimator method or scanning beamformer, respectively. Local approaches to solving the inverse problem currently provide the best performance and spatial resolution, but their application is limited by the availability of synchronized neural sources. In [Kuznetsova et al. \(2021\)](#), we propose an approach that enables an adaptive beamformer operating in an environment with highly synchronized sources to regain performance.

In addition to the spatial characteristics determined by the geometrical properties of electrically active neuronal populations, the data recorded using electrophysiological methods with high temporal resolution also have dynamic characteristics reflecting the frequency-time properties of neuronal source activity. Due to energetic reasons and simultaneous requirements for stability and high response rate of the whole system, neuronal populations tend to switch between states of excitation and inhibition, which generates rhythmic components of electrical activity of the brain recorded both invasively and noninvasively [Buzsaki \(2006\)](#); [Buzsáki et al. \(2012\)](#). The brain is a system consisting of a large number of low-level functionally specialized zones. To fulfill its mission such a distributed system must support selective and context-dependent information exchange between its various elements - functionally specialized neuronal populations [Lachaux et al. \(1999\)](#). This is how the fundamental principle of functional integration [Friston \(2002\)](#), which underlies all brain function [Rizzolatti et al. \(2018\)](#), is implemented.

According to one hypothesis, information exchange between neuronal ensembles is carried out by dynamic (time-varying) mutual synchronization of sequences of their excitation//inhibition states. When observing the activity of relatively large populations, such switching between excitation//inhibition states is reflected in the presence of oscillations, whose degree of synchronization can be estimated using the coherence function. The presence of intervals of the elevated statistically reliable coherence between oscillations generated by different neuronal ensembles can indicate the ongoing processes of information exchange. Such an idea of organization of effective information transmission channels between neuronal ensembles due to synchronization was named in the literature as "communication through coherence"(CTC) [Fries \(2015\)](#). In a nutshell, synchronization of oscillations reflects the processes of dynamic communication in a network of functionally specific brain areas. One type of functional

connectivity manifests itself as the presence of statistically significant sustained phase difference of the rhythmic activity of a pair of neuronal populations. As a rule, the presence of a non-zero phase delay makes it possible to draw conclusions about the causality of such interaction and assess the direction of information distribution, identify the master and slave neuronal populations. However, two populations of neurons are often connected by bidirectional connections, which leads to near-zero phase delay in their [Rajagovindan and Ding \(2008\)](#) oscillations. Also, in the case of the coupling between populations implemented via high-speed myelinated fibers, the phase delay can be limited to a small fraction of the period of their rhythmic activity. The processes of rearrangement of the populations' rhythm and tuning to a common rhythm are also accompanied by a near-zero lag between the two signals [Pikovsky et al. \(2001\)](#); [Schuster and Wagner \(1989\)](#). Finally, the dependence of the activity of the two populations on the third, also leads to scenarios of a functional relationship with zero phase difference. As invasive measurements show, it is the interactions characterized by a small phase angle that are most frequently observed in the experimental data [Roelfsema et al. \(1997\)](#); [Singer \(1999\)](#); [Engel A.K. \(2001\)](#).

Due to purely physical electromagnetic properties of the head tissue and characteristic frequencies of activity of neuronal populations, signal propagation from neuronal populations to the sensors is almost instantaneous [Hamalainen et al. \(1993\)](#). **So volume conduction complicates the detection of exactly the most frequent neural networks whose functional connectivity is characterized by zero or small phase delay.** This is due to the fact that, when projected onto sensors, the effect of such true instantaneous physiological and functional connectivity in the activity of neuronal populations appears indistinguishable from the effect of volumetric conductivity caused exclusively by the physical properties of the head as a conductor or the properties of the magnetic field [Stam et al. \(2007\)](#). **In [Ossadtchi et al. \(2018\)](#) we proposed a method by which we managed for the first time to isolate from multichannel MEG measurements functional networks of truly synchronized with small phase delay sources.**

As a rule, a large number of subjects take part in cognitive experiments, and the results of solving the inverse problem obtained by the methods described above are averaged over the subjects. As the main result the researchers typically use such maps of cortically distributed activity or neural networks whose activation, on average in the population, accompanies the performance of the cognitive or motor task under study [Papanicolaou \(1998\)](#).

Typically, the difference in the cortical activity observed between two experimental conditions is of interest and is attributed to the cognitive process under study [Poldrack \(2018\)](#). Averaging across subjects involves alignment of individual cortical envelopes with some canonical cortical surface using methods of spatial transformation (warping) to support further visualization and averaging of the inverse solutions. Such approach is realized in the overwhelming number of laboratories of the world and in fact is a standard of carrying out of group-level neuroimaging studies. However, it has significant disadvantages and inefficiently uses the information contained in the measured data [Larson et al. \(2014\)](#).

It is reasonable to take into account the fact that fuzzy similarities of cortical activity profiles across subjects coexist with individual, person-specific cortical activations. **Based on this assumption, it is possible to construct an iterative algorithm**

that would provide a solution to the inverse problem operating with whole group sensor data. In [Kozunov and Ossadtchi \(2015\)](#) we use these fuzzy inter subject similarities of activity extracted from the data to reduce uncertainty in solving the inverse problem.

The differential activation map obtained by solving the inverse problem serves as a kind of distance between the spatiotemporal dynamics of brain activity observed in a pair of experimental conditions and correlated (only in the case of a correctly designed experiment) with the cognitive process under study. It is clear, however, that such an approach reveals only very rough differences between the profiles of neuronal activity, reduced simply to the difference in the intensity of neuronal populations and essentially depending on the degree of their phase binding to the moment when the subject starts processing the presented stimulus.

The new turn in the development of machine learning technology has given researchers from various scientific fields the opportunity to experiment with flexibly architected neural networks, implementing custom cost functions and applicable directly to the measured signals, thus bypassing the traditional feature extraction step typical for the earlier stage of machine learning technology development. In contrast to the classical machine learning, based on manual determination of informative features, deep learning techniques allow us to automate this process and perform extraction of features with the first few layers of the neural network, wired in accordance to a particular field of knowledge and focused on the adaptive implementation of feature extraction methods, accepted in a particular field of science [Elmarakeby et al. \(2021\)](#); [Petrosyan et al. \(2020a\)](#). Thus, the problem of determining the distance between the spatiotemporal dynamics of brain activity in two conditions can be formulated as the task of constructing a classifier of brain states based on the measured activity. Given that the neural network-based solver configured for this purpose is not restricted to the implementation of a linear mapping, we can expect a more comprehensive reflection of the characteristic properties of neuronal activity than that furnished by the traditional techniques based on the differential activation. Using such a solver we can introduce a distance, between two experimental conditions, proportional to the probability of their correct classification. The higher the probability, the more distant the two conditions are in some learned neural network feature space. In some cases, using specific formulations of the cost function, it is possible to take into account the topology of the corresponding feature space [Sabbagh et al. \(2019\)](#).

In the case of a network architecture that allows the interpretation of weights of the first layers in accordance with the physiological knowledge and reflecting the physical principles of a particular neuroimaging modality, it is possible, by analyzing the weights of the neural network, to solve the problem of localizing pivotal to the downstream task neural populations and identifying the dynamic aspects of their activity [Petrosyan et al. \(2020a\)](#). Thus, the use of interpretable neural network architectures, coordinated with the existing knowledge in the specific field of science, also allows to implement the process of automatic knowledge extraction from the cognitive experiments data organized according to the classical scheme and contrasting the neural activity between experimental conditions that differ, ideally, only in the presence of the studied cognitive process in one of the conditions.

Initially, the technology of brain-computer interfaces implied more obvious and traditional application of machine learning methods and pattern recognition techniques to processing brain activity signals. Modern neural network architectures, that have already found application in this field, have allowed us to create prototypes of systems for restoration of motor and speech functions by decoding neural activity in commands to prosthetic limbs, speech synthesis device or a system for displaying text messages on the screen. However, the "greed" of machine learning algorithms based on modern architectures with a large number of parameters, often leads them to using information coming not from the brain activity, but from other processes, accompanying the process of command generation. The most prominent example of sources of such information is the electrical activity generated by changes in the tone of the scalp muscles, oculomotor muscles, neck muscles, and tongue muscles. **Use of interpretable architectures followed by analysis of geometrical and frequency properties of the most informative sources is capable of giving an answer about the nature of signals used by the trained classifier or decoder.**

Traditionally, and in accordance with their purpose, brain-computer interface systems (BCI) imply real-time operation, decoding brain activity registered in the time window immediately preceding the current moment of time. As a rule, the rhythmic components of brain activity are informative in such systems. To ensure naturalness of use, it is necessary to reduce the delay between the moment of generation of the brain state corresponding to a certain command and the moment of time when the given command appeared to be decoded. Especially relevant is such a reduction of delay for bidirectional interface systems, that involve not only decoding, but also generating the feedback via direct cortical or peripheral stimulation.

Another subdomain where feedback delay reduction appeared relevant is that of Neurofeedback. Neurofeedback technology involves visualization of a certain aspect of the user's brain activity followed by the development of skills to control that activity in an arbitrary manner. As a rule, the amplitude of a certain brain rhythm (alpha, beta, theta) [Buzsaki \(2006\)](#) is used as a feedback parameter. Brain rhythmic activity tends to occur in bursts of relatively short length, on the order of 200-300 ms, comparable to the delay in the signal processing pathway of the vast majority of commercial and laboratory neurofeedback systems, which is 500 - 1000 ms and higher. Thus, it appears that the feedback is presented to the user with a long delay with respect to the reinforced event (burst of rhythmic activity), which leads to a significant decrease in the efficiency of the training process in the neurofeedback paradigm [Ossadtchi et al. \(2017a\)](#); [Belinskaia et al. \(2020a\)](#).

Such a delay in the signal processing pipeline designed to extract brain's rhythmic activity parameters consists of two parts. The first part is related to the Heisenberg-Gabor uncertainty, which imposes, in the absence of additional information, a fundamental limitation on the joint accuracy of determining the frequency of a periodic signal and the timing of its occurrence. The second component of the delay is related to purely technical issues depending on the communication protocols between the electroencephalograph and the computer, as well as the internal processes of the computer operating system used to process the EEG signal and generate the feedback signal. **The development of special methods for narrow-band filtering using additional information about the dynamic properties of the target signal, as well as the implementation of the signal processing algorithms in the real-**

time operating system deployed on board of the electroencephalograph, can significantly reduce the overall delay in the feedback signal presentation and increase the efficiency of training in the neurofeedback paradigm [Belinskaia et al. \(2020a\)](#).

## 3 New methods for solving the EEG and MEG inverse problem

### 3.1 Adaptive beamformer with modified covariance matrix

#### 3.1.1 Introduction

The essence of the inverse problem (IP) of the MEG is to convert the signals of the magnetic field sensors into activation maps of the cerebral cortex. Thus, the spatial resolution of MEG, as a technology for mapping neuronal sources on the cerebral cortex, critically depends on the approach used to solve the inverse problem. The minimum norm method (MNN, MNE) is currently the most commonly used approach, the stability of which is achieved at the expense of the low spatial resolution of the resulting cortical maps [Hincapié et al. \(2016\)](#).

In recent years, local methods for solving IP based on the use of adaptive beamformers or linearly constrained minimum variance beamformers (LCMV BF) [Van Veen et al. \(1997\)](#); [Sekihara et al. \(2001\)](#); [Greenblatt et al. \(2005b\)](#) have gained popularity. When operating in the environment with a small number of *uncorrelated* sources, LCMV BF demonstrate excellent performance as neuronal activity detectors and provide high spatial resolution [Jaiswal et al. \(2020\)](#). However, LCMV BFs fail when there are sources with correlated time series. This leads to a significant decrease in the signal-to-noise ratio in the time series obtained using BFs [Sekihara and Nagarajan \(2008\)](#); [Kuznetsova et al. \(2021\)](#) and meaningless maps of power distribution over the cerebral cortex. This situation precludes the wider use of this potentially promising approach to solving the inverse problem, especially since the fundamental mechanisms of brain functioning that implement the principle of functional integration [Fries \(2015\)](#) naturally imply the presence of sources with correlated activity, as described in **Introduction**.

To solve this problem, we have developed a new approach that involves modifying the covariance matrix of the observed data [Kuznetsova et al. \(2021\)](#). Our method is based on the projection operation applied to the vectorized covariance matrix of sensor signals. Such a projection, applied in the product-space of sensor data covariance matrices, whose dimension is equal to the square of the number of sensors, does not remove the activity of correlated sources from the data and, accordingly, from the covariance matrix of the sensor space. Instead, it separates the additive contribution to the covariance structure explained by the power of sources from the contribution modulated by the source timeseries correlation. Retention of the former component creates a fairly accurate approximation of the ideal covariance matrix of sensor data, which could hypothetically be observed in the case of uncorrelated source activity. Next, we use this modified covariance matrix to calculate the weights of the spatial filters in accordance with the standard approach to calculating the LCMV BF weights.

### 3.1.2 Data Model and Problem Statement

The vast majority of methods for solving the EEG and MEG inverse problem are based on a linear model of observing the signal vector  $\mathbf{x}(t) = [x_1(t), \dots, x_M]^T$  by an array of  $M$  sensors at a time  $t$  associated with the source activity signals  $s_i(t)$ ,  $i = 1, \dots, R$  by the following observation equation:

$$\mathbf{x}(t) = \sum_{i=1}^R \mathbf{g}_i s_i(t) + \mathbf{n}(t) = \mathbf{G} \mathbf{s}(t) + \mathbf{n}(t). \quad (1)$$

where  $\mathbf{s}(t) = [s_1(t), \dots, s_R(t)]^T$  is a  $R$ -dimensional vector  $[R \times 1]$  of  $R$  source signals. We assume here that the evoked activity  $\mathbf{x}(t)$  is generated by a relatively small number  $R$  of focal cortical sources. Here  $\mathbf{g}_i$  is the topography of the  $i$ -th equivalent current dipole approximating the neural source, and  $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_R]$  is the matrix of oriented source topographies. The noise term  $\mathbf{n}(t)$  is the sum of the residuals of induced and task unrelated activity that is assumed to be sufficiently suppressed by the event-related averaging procedure, resulting in a high signal-to-noise ratio (SNR) of target signals  $s_i(t)$  in the observed mixture  $\mathbf{x}(t)$ .

The beamformer method developed here formally does not require the additive noise  $\mathbf{n}(t)$  (Van Veen et al. (1997)) to be spatially uncorrelated, and therefore we make no assumptions about its spatial covariance structure. It is assumed, however, without loss of generality, that the covariance matrix of the observed data  $\mathbf{C}_x = E\{\mathbf{x}(t)\mathbf{x}^T(t)\}$  has full rank. In the case when this is not the true, it is possible to reduce the dimension of the sensor space and perform all operations in a space of lower dimension.

The locations of the  $R$  active sources are unknown, and finding them is the main goal of solving the EEG and MEG inverse problem. Traditionally, we approach this problem with known forward model that maps each possible  $j$ -th location of a dipolar neural source to the topography vector  $\mathbf{g}_j$ . It is important to understand that while we know the vector  $\mathbf{g}_j$  for each point of the discretized surface of the cerebral cortex, and there are usually several tens of thousands of such points, we do not know which  $R$  of these points contain active sources whose the signals  $s_j(t)$  form the observed mixture  $\mathbf{x}(t)$ . Moreover, the number  $R$  of such sources is also unknown to us. Therefore, one of the methods for solving the inverse problem implies scanning with LCMV BF over all possible cortical mantle locations. Within the framework of this approach, for each potential  $j$ -th location, the BF is constructed, the output power is estimated, and a power distribution map is formed for the entire cortex comprising the power of each individual node. Despite the fact that the estimates of the location of sources obtained as locations of local maxima of such a map are biased (Greenblatt et al. (2005a)), this approach in practice and in the absence of pronounced synchronous sources provides a very acceptable quality for solving the inverse problem.

The LCMV BF (Van Veen et al. (1997); Sekihara et al. (2001); Greenblatt et al. (2005b)) is a method for computing a spatial filter vector  $\mathbf{v}$  that provides a minimum output signal power  $s(t) = \mathbf{v}^T \mathbf{x}(t)$  constrained by the requirement of unit signal transmission coefficient from the “direction” in which the target signal is located and which is determined by the corresponding forward model vector  $\mathbf{g}_i$ . The logic of this approach is that the requirement for a minimum power at the output achieves detuning from noise sources, and the restriction on unity gain guarantees, in the absence of

correlated sources, that the signal of the target source will not be suppressed and will be present at the output without changes. Without restriction on the unit transfer coefficient, the corresponding optimization problem obviously has the trivial solution  $\mathbf{v} = \mathbf{0}$ . In other words, for each location with index  $i$ , the following optimization problem is solved on the discretized model of the cerebral cortex [Sekihara et al. \(2001\)](#):

$$\begin{aligned} & \underset{\mathbf{v}_i}{\text{minimize}} && \mathbf{v}_i^T \mathbf{C}_x \mathbf{v}_i \\ & \text{subject to} && \mathbf{v}_i^T \mathbf{g}_i = 1, \end{aligned} \quad (2)$$

where  $(\cdot)^T$  is the transposition operator,  $\mathbf{C}_x = E\{\mathbf{x}(t)\mathbf{x}(t)^T\}$  is the covariance matrix of sensor data, and the vector  $\mathbf{g}_i = \mathbf{g}(\mathbf{r}_i)$  is the topography of the  $i$ -th source located at the point with coordinates  $\mathbf{r}_i$ . Using the method of Lagrange multipliers, one can find the following analytical solution

$$\mathbf{v}_i^T = [\mathbf{g}_i^T \mathbf{C}_x^{-1} \mathbf{g}_i]^{-1} \mathbf{g}_i^T \mathbf{C}_x^{-1} \quad (3)$$

The computed spatial filters  $\mathbf{v}_i$  can then be used to reconstruct the original time series vector estimates as

$$\widehat{\mathbf{s}}_i(t) = \mathbf{v}_i^T \mathbf{x}(t). \quad (4)$$

It is also possible to use the beamformer in scan mode and calculate the power distribution profile  $\sigma_i^2 = \text{Var}(\mathbf{r}_i)$  for the entire set of  $N$  cortical locations  $\mathbf{r}_i$ ,  $i = 1, \dots, N$  which can be done without explicit calculation of  $\mathbf{V}_i$ :

$$\sigma_i^2 = \mathbf{v}_i^T \mathbf{C}_x \mathbf{v}_i = [\mathbf{g}_i^T \mathbf{C}_x^{-1} \mathbf{g}_i]^{-1}. \quad (5)$$

As follows from the equations (3) and (5), given a fixed direct model  $\mathbf{g}_i$ , the covariance matrix completely determines the weights of the AFL and the output power of the source estimates when applying to the data  $\mathbf{x}(t)$ .

The described approach does not introduce any assumptions about the number of active sources or their spatial distribution. However, LCMV BF is known to provide spatial superresolution only when the measured neuronal activity is generated by a small number of focal uncorrelated cortical sources ([Borgiotti and Kaplan \(1979\)](#)) and, in the general case, gives biased estimates of active source locations [Greenblatt et al. \(2005b\)](#).

The beamformer output power in the case of two correlated sources is expressed as ([Sekihara and Nagarajan \(2008\)](#)):

$$\widehat{\sigma}_i^2 = 1 - (\rho_{12}^{ss})^2, \quad i = 1, 2 \quad (6)$$

Thus, in the case of full synchronization, the output of the adaptive beamformer is simply zero. Intuitively, this can be understood as follows. To meet the constraint, the BF must ensure unity target source gain, and the adaptive beamformer-optimized functionality requires output power to be minimized. In the presence of another source with correlated activity, the beamformer adjusts the weight vector so that, on the one hand, the unity gain constraint on the given "direction" is observed, and, on the other hand, the activity of the correlated sources is subtracted from the target activity to minimize the output power. Therefore, in the case of perfect correlation, the beamformer

produces zero SNR with respect to the activity of the target source, since it manages to choose such weights with which the activity of one source completely compensates for the activity of another. Indeed, as long as the topographies ( $\mathbf{g}_i, \mathbf{g}_j$ ) of these two synchronous sources are non-orthogonal, the requirement for unity gain of the target source signal ( $\mathbf{g}_i$ ) can be provided by simple scaling of the vector of the BF coefficients.

The situation is exacerbated in an environment with a large number of correlated sources. Even a moderate correlation of the time series of the source with the activity of other populations leads to a significant decrease in the signal-to-noise ratio at the LCMV BF output. Consider a medium with three active sources of unit variance with time series  $s_1(t), s_2(t), s_3(t)$ , characterized by pairwise correlation coefficients  $\rho_{12}^{ss}, \rho_{13}^{ss}, \rho_{23}^{ss}$ .

It is easy to show that in this case the BF output power  $\hat{\sigma}_i, i = 1, 2, 3$  depends on the pairwise correlation coefficients  $\rho_{ij}^{ss}$  as

$$\hat{\sigma}_1^2 \propto \frac{F}{(\rho_{23}^{ss})^2 - 1}, \quad \hat{\sigma}_2^2 \propto \frac{F}{(\rho_{13}^{ss})^2 - 1}, \quad \hat{\sigma}_3^2 \propto \frac{F}{(\rho_{12}^{ss})^2 - 1}, \quad (7)$$

where

$$\text{where } F = (\rho_{12}^{ss})^2 + (\rho_{13}^{ss})^2 + (\rho_{23}^{ss})^2 - 2\rho_{12}^{ss}\rho_{13}^{ss}\rho_{23}^{ss} - 1.$$

Panels A and B of 1 show color-coded power estimates of the three sources  $s_1, s_2$  and  $s_3$  for different degrees of correlation between the first source and the other two sources, when  $s_2$  and  $s_3$  are a strongly connected pair and  $\rho_{23} = 0.9$ . As expected, the output power of the second and third sources is reduced primarily due to their strong mutual coupling, figure 1.B. In this case, the low power of the  $s_2$  and  $s_3$  estimates obtained complicates their detection and makes it problematic to implement the strategies proposed in (Dalal et al. (2006), Popescu et al. (2008)) to eliminate problems caused by source correlation. In addition, we also see a rapid decline in the  $s_1$  cardinality estimate as the relationship between this first source and either of the two remaining sources grows, figure 1.A.

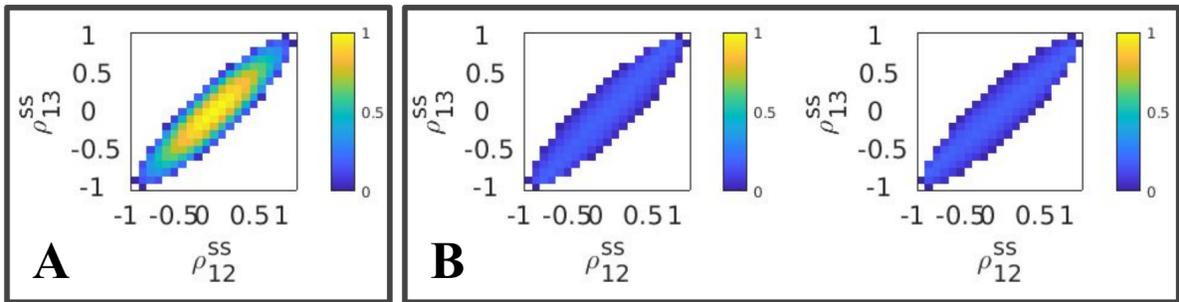


Рис. 1: Power (variance) at the LCMV BF output in the case of three active sources with pairwise correlations  $\rho_{12}^{ss}, \rho_{13}^{ss}$  and  $\rho_{23}^{ss} = 0.9$ . Estimated variance  $\hat{\sigma}_i, i = 1, 2, 3$  for each of the three sources depending on pairwise correlation  $\rho_{12}^{ss}, \rho_{13}^{ss}$  is color coded. In each graph, the axis reflects the relationship of the first source  $s_1$  with the other two sources  $s_2$  and  $s_3$ , quantified by the correlation coefficients  $\rho_{12}^{ss}$  and  $\rho_{13}^{ss}$ . Impossible combinations of correlation coefficient values for given sources of unit variance are shown in white.

### 3.1.3 New method

$M \times M$  the covariance matrix of sensor signals  $\mathbf{C}_x = E\{\mathbf{x}(t)\mathbf{x}^T(t)\}$  plays a key role in the calculation of the LCMV BF coefficients. Using the observation model of sensor signals  $\mathbf{x}(t)$  and the definition of the covariance matrix, assuming zero mean of the signals, the vectorized covariance matrix of sensor signals  $vec(\mathbf{C}_x)$  can be expressed in terms of the elements  $c_{ij}^{ss}$ ,  $i, j = 1, \dots, R$  of the covariance matrix of the original space  $R$  of active sources and their topographies  $\mathbf{g}_i, \mathbf{g}_j$  as

$$vec(\mathbf{C}_x) = vec(E\{\mathbf{x}(t)\mathbf{x}^T(t)\}) = \underbrace{\sum_{i=1}^R vec(\mathbf{g}_i\mathbf{g}_i^T)}_{\text{green box}} c_{ii}^{ss} + \underbrace{\sum_{i=1}^R \sum_{j=i+1}^R vec(\mathbf{g}_i\mathbf{g}_j^T + \mathbf{g}_j\mathbf{g}_i^T)}_{\text{red box}} c_{ij}^{ss} + vec(\mathbf{C}_n) \quad (8)$$

where  $\mathbf{C}_n$  - noise covariance matrix and elements  $\mathbf{q}_{ij} = vec(\mathbf{g}_i\mathbf{g}_j^T)$  we will call 2-topographies.

It is important to understand that in practice the expectation operator  $E\{\}$  used in this expression is approximated by averaging the external product of sensor signals  $\mathbf{x}(t)$  over the moments of presentation of a stimulus (image, sound) to the test subject or the beginning of the motor task, Usually, in cognitive experiments, about 100 such presentations are performed, which, as a rule, is sufficient to form an informative estimate of the average external product of the sensor signal vector.

The (8) equation demonstrates that the matrix  $\mathbf{C}_x$  can be decomposed into two types of additive components: 1) auto-components modulated by power  $c_{ii}^{ss}$  active sources  $\sum_{i=1}^R vec(\mathbf{g}_i\mathbf{g}_i^T)c_{ii}^{ss}$ , 2) cross components, whose contribution to the  $\mathbf{C}_x$  structure is modulated by the degree of activity correlation of the corresponding pairs of sources  $\sum_{i=1}^R \sum_{j=i+1}^R (\mathbf{g}_i\mathbf{g}_j^T + \mathbf{g}_j\mathbf{g}_i^T)c_{ij}^{ss}$ . It is obvious that these cross terms are present in the covariance matrix due to non-zero off-diagonal elements  $c_{ij}$  of the neural source covariance matrix. It is the presence of non-zero off-diagonal elements  $c_{ij}$  of the covariance matrix of neural sources that leads to a decrease in the performance of the adaptive beamformers.

To reduce the contribution of non-zero off-diagonal elements of the source covariance matrix to the sensor signal covariance matrix, we propose to use a specially constructed projection matrix that operates in  $M^2$ -dimensional space and minimizes the contribution of cross components to  $\mathbf{C}_x$ . The corresponding processing pipeline is shown in 2. We apply this projection matrix  $\mathbf{P}$  to the vectorized covariance matrix of the data to dampen the contribution of the cross terms. The resulting matrix after projecting the shape back into a square matrix and performing the [Duin and Pękalska \(2010\)](#) spectral reflection operation to ensure the positive definiteness of the resulting matrix by replacing a small number of negative eigenvalues small in absolute value by their absolute values, approximates the covariance matrix of sensor signals, which could would be obtained in the absence of a correlation between the activity of the sources.

Further, the transformed matrix  $\tilde{\mathbf{C}}_x^{abs}$  is used to calculate the weights or directly the power at the BF output in accordance with the standard expressions 3 and 5

### 3.1.4 Calculation and application of the projection operator

We have proposed two different procedures for calculation of the projection operator. The first one is based on the singular value decomposition (SVD) of the 2-topography

1. Vectorize sensor-space covariance matrix

$$\mathbf{C}_x = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1M} \\ \vdots & \vdots & \vdots & \vdots \\ c_{M1} & c_{M2} & \dots & c_{MM} \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{12} \\ \vdots \\ c_{1M} \\ \vdots \\ c_{MM} \end{bmatrix}_{[M^2 \times 1]}$$

The diagram illustrates the decomposition of the vectorized covariance matrix  $\mathbf{C}_x$  (red dot) into two components: a green vector representing the sum of auto-terms  $\sum_{i=1}^N \text{vec}(\mathbf{g}_i \mathbf{g}_i^T) c_{ii}^{ss}$  and a black vector representing the sum of pairwise cross products  $\sum_{i=1}^N \sum_{j=i+1}^N \text{vec}(\mathbf{g}_i \mathbf{g}_j^T + \mathbf{g}_j \mathbf{g}_i^T) c_{ij}^{ss}$ .

2. Project and keep within PDM manifold

$$\tilde{\mathbf{C}}_x = \text{vec}^{-1} \left( \begin{bmatrix} \tilde{c}_{11} \\ \tilde{c}_{12} \\ \vdots \\ \tilde{c}_{1M} \\ \vdots \\ \tilde{c}_{MM} \end{bmatrix} \right) = \text{vec}^{-1} \left( \mathbf{P} \cdot \begin{bmatrix} c_{11} \\ c_{12} \\ \vdots \\ c_{1M} \\ \vdots \\ c_{MM} \end{bmatrix} \right)$$

The diagram shows the projection of the vectorized covariance matrix  $\mathbf{C}_x$  (red dot) onto a manifold. The projection result is a red dot labeled  $\tilde{\mathbf{C}}_x$ . A dashed line indicates the projection path.

3. Apply usual beamformer with projected matrix

$$\begin{aligned} & \underset{\mathbf{w}_i}{\text{minimize}} \quad \text{Tr}\{\mathbf{w}_i^T \tilde{\mathbf{C}}_x \mathbf{w}_i\} \\ & \text{subject to} \quad \mathbf{w}_i^T \mathbf{g}_i = \mathbf{I} \end{aligned}$$

Рис. 2: The main stages of the proposed approach: 1. Consider the vectorized covariance matrix of sensor signals  $\mathbf{C}_x$ , shown by the red dot, as an element of the space  $M^2$ . It can be decomposed into a sum of two non-orthogonal vectors: auto-terms corresponding to the power of the source  $\sum_{i=1}^R \text{vec}(\mathbf{g}_i \mathbf{g}_i^T) c_{ii}^{ss}$  and pairwise cross products of the original topographies weighted with source activity covariance coefficients  $\sum_{i=1}^R \sum_{j=i+1}^R (\mathbf{g}_i \mathbf{g}_j^T + \mathbf{g}_j \mathbf{g}_i^T) c_{ij}^{ss}$ . 2. The precomputed projection matrix  $\mathbf{P}$  is applied to the matrix  $\mathbf{C}_x$ , then measures are taken to return the projection result back to a variety of positive definite matrices, and then the result is converted back to  $M \times M$  matrix  $\tilde{\mathbf{C}}_x^{abs}$ . 3. Finally, the adaptive beamformer spatial filters are calculated as usual, but using the modified sensor space data covariance matrix.

matrix  $\mathbf{G}_{pwr} = \mathbf{q}_{ii}$ ,  $i = 1, \dots, N$ . The SVD allows one to estimate the subspace of  $M^2$  dimensional space of 2-topographies, in which the main variability of the columns of the matrix  $\mathbf{G}_{pwr}$  is concentrated. Considering that the columns of the matrix  $\mathbf{G}_{pwr}$  are auto-2-topographies, that is, 2-topographies with the same indices, and their contribution to the covariance matrix is modulated exclusively by the power of the corresponding sources, we will call this a subspace of power  $\mathcal{S}_{pwr}^K$  and define it as a linear span spanned by the first  $K$  left eigenvectors of the matrix  $\mathbf{G}_{pwr}$  corresponding to the  $K$  maximum eigenvalues. Further, the projection operator is calculated as the projection operator of the observed covariance matrix into the subspace  $\mathcal{S}_{pwr}^K$ :

$$\mathbf{G}_{pwr} = \mathbf{U}_{pwr} \mathbf{S}_{pwr} \mathbf{V}_{pwr}^T \mathbf{P}_{pwr} = \mathbf{U}_{pwr}^{1:K} \mathbf{U}_{pwr}^{1:K T} \quad (9)$$

Thus, by applying  $\mathbf{P}_{pwr}$  to the vectorized covariance matrix  $\mathbf{C}_x$  we emphasize the components of  $\mathbf{C}_x$  which are due to the power, not the covariances of the neural sources. To retain the positive definiteness property in the projection matrix we apply the spectral reflection procedure [Duin and Pękalska \(2010\)](#). As a rule, the contribution of eigendirections modulated by negative eigenvalues is rather small and does not exceed 10-20% of the total energy in the eigenvalue spectrum of the projected matrix.

The procedure for choosing the optimal projection  $K$  rank is described in detail in [Kuznetsova et al. \(2021\)](#) and consists in finding such a maximum value of the projection rank for which the weakening of the contribution of the components from  $\mathcal{S}_{cor}^K$  exceeds the weakening of the contribution from  $\mathcal{S}_{pwr}^K$ , which is inevitable due to the non-orthogonality of the two subspaces.

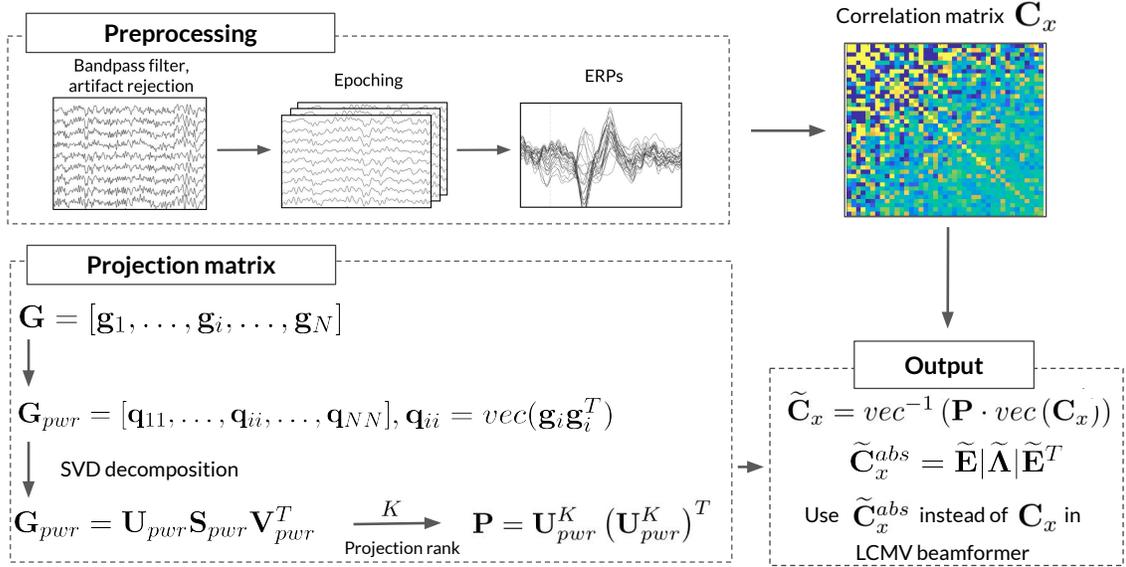


Рис. 3: Schematic representation of the ReciPSIICOS algorithm. After preprocessing, the evoked response (ERP) and its covariance matrix  $\mathbf{C}_x$  are calculated. The projection matrix is calculated and applied to the vectorized  $\mathbf{C}_x$ , after which the projection result is returned to the square form. To give the result of the projection the property of the covariance matrix, which consists in positive definiteness, we apply the procedure of spectral reflection. Further, the transformed matrix  $\tilde{\mathbf{C}}_x^{abs}$  is used to calculate the weights or directly the power at the BF output in accordance with the standard expressions 3 and 5

Also, in [Kuznetsova et al. \(2021\)](#) one more projection is proposed, whose calculation is based on the projection into the space orthogonal to the subspace  $\mathcal{S}_{cor}^K$ , which is defined and operationalized similarly to  $\mathcal{S}_{pwr}^K$ , but based on the matrix  $\mathbf{G}_{cor}$ , whose columns represent cross 2-topographies  $\mathbf{q}_{ij} = \text{vec}(\mathbf{g}_i \mathbf{g}_j^T)$ ,  $i \neq j$ . However, taking into account that the variability of the columns is also modulated by the norms of the corresponding 2-topographies, in order to preserve the contribution from  $\mathcal{S}_{pwr}^K$  after the projection, it is necessary to perform the SVD of the matrix in space where

matrix  $\mathbf{C}_{pwr} = \mathbf{G}_{pwr} \mathbf{G}_{pwr}^T$  is diagonalized. This approach has been dubbed Whitened ReciPSIICOS.

### 3.1.5 Results

To evaluate the performance of the new approach we first performed a set of realistic simulations as described in [Kuznetsova et al. \(2021\)](#). We conducted a comparative analysis of the new algorithm's performance and matched it against the classical LCMV BF, as well as against the global approach using the MNE method, the most commonly used algorithm for solving the EEG and MEG inverse problem.

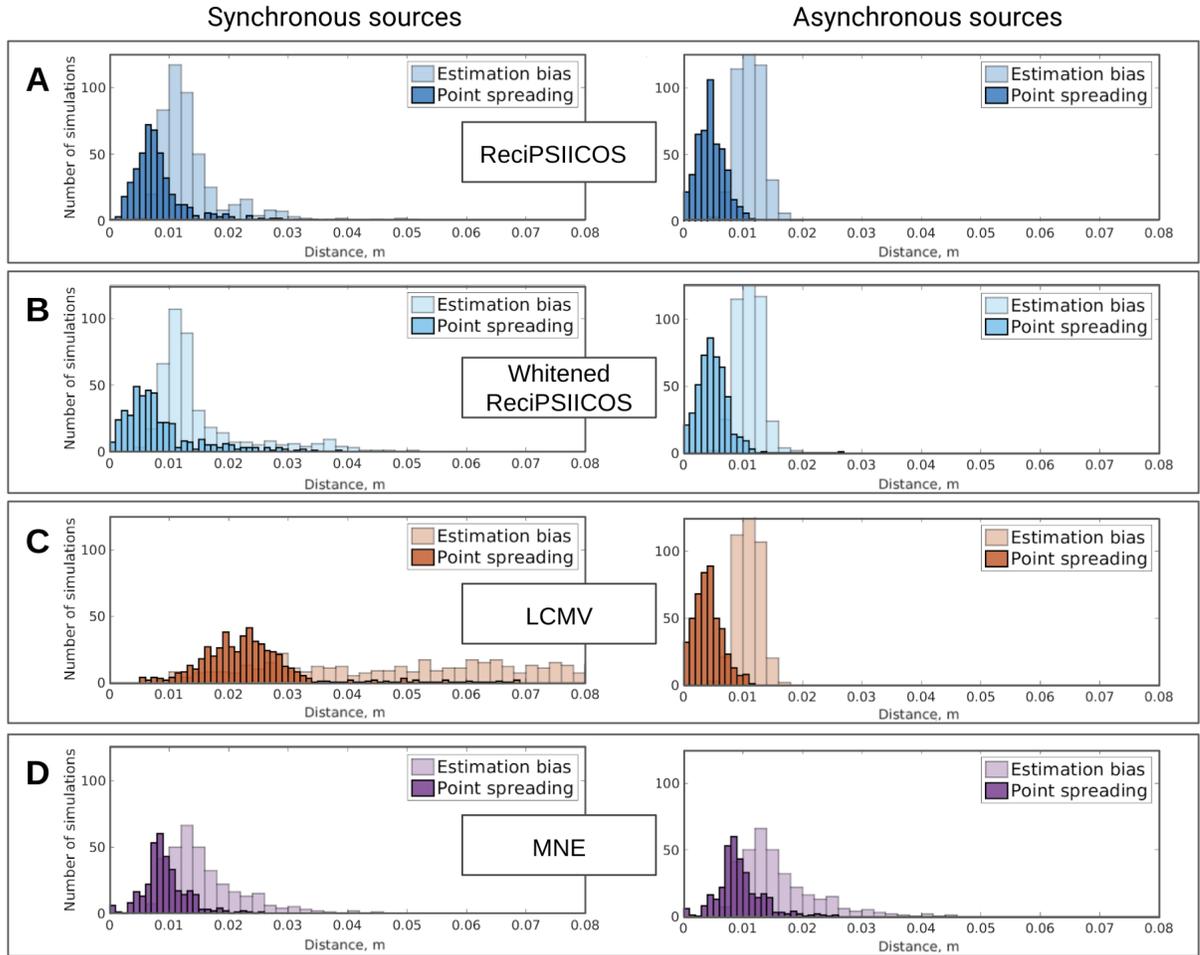


FIG. 4: Distribution of bias and scatter of localization result for 500 Monte Carlo iterations,  $\text{SNR} = 4$ . Results are computed for ReciPSIICOS (A), Whitened ReciPSIICOS (B), LCMV BF (C), and MNE (D).

As the accuracy criteria we used the shift and spread that point sources suffer when they are first projected on the sensors with the ideal forward operator and then recovered using the specific inverse solver based on the perturbed forward model to avoid the 'inverse crime'. Comparative results of 4 methods are shown in Figure 4. As can be seen from the histograms obtained in the framework of multiple Monte Carlo

iterations, characterized by a random choice of the location of several neural sources, Whitened ReciPSIICOS and ReciPSIICOS provide a smaller spread compared to the classical LCMV BF and MNE.

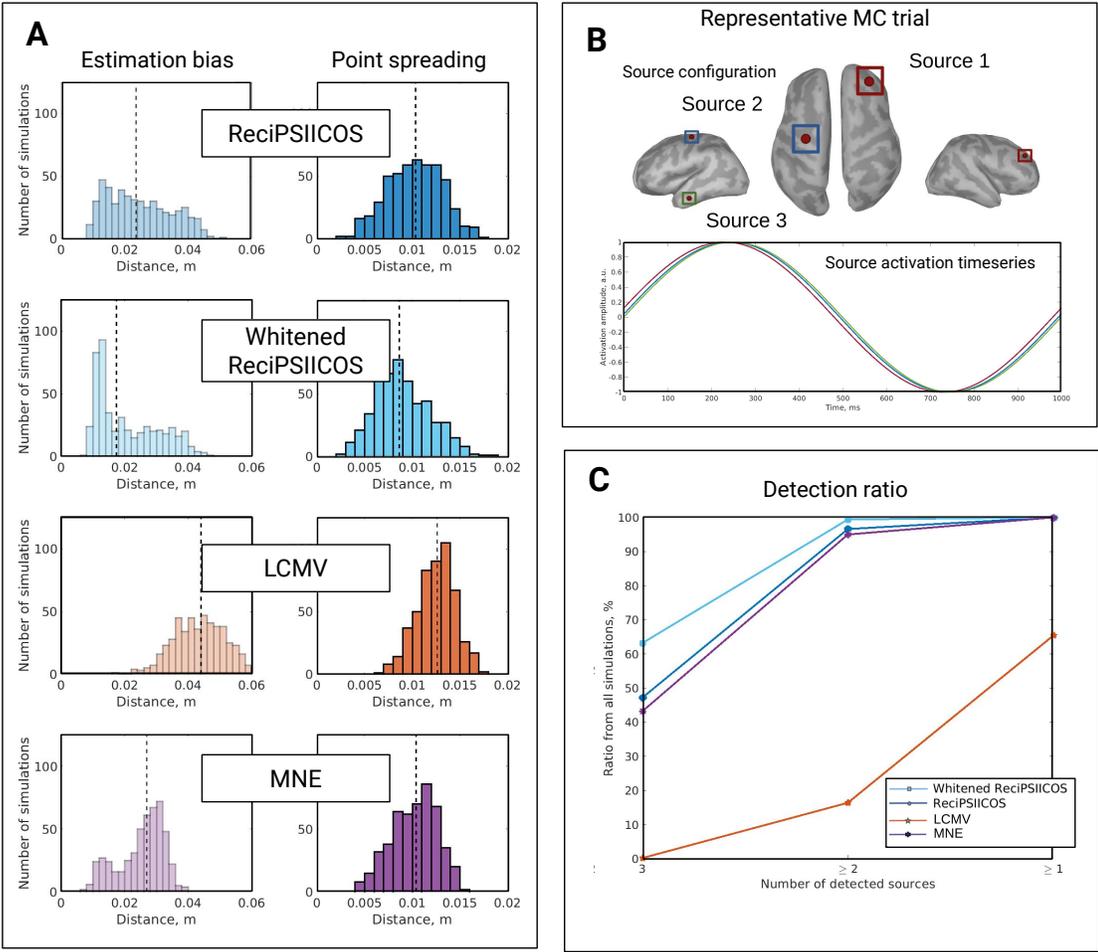


Рис. 5: Modeling of three synchronously active sources. **A**. Distribution of two reconstruction quality measures: estimation bias and point scatter values for the four source estimation methods: ReciPSIICOS, Whitened ReciPSIICOS, LCMV BF, and MNE. **B**. One representative Monte Carlo study: random selection and activation of three cortical sources with synchronous sinusoidal activity profiles with average phase difference of  $\pi/3$  and random jitter. **C**. Distribution of the number of detected sources for all iterations.

In addition, in the case when activation profiles of all three sources are moderately correlated ( $r = 0.6$ ) within each of the three pairs, the proposed approach provides significantly better performance in the problem of detecting these sources and has the smallest bias and scatter. Figure 5 shows the comparative performance of the four methods. In particular, 5.C shows the probabilities of correct detection of one, two

and all three simulated sources by four different methods. It can be seen that both modifications of the new approach detect all three active sources in more than 80% of cases. While the MNE, which is the closest in quality, does so in only 40% of cases.

We also applied our approach to MEG datasets from two experiments involving two different auditory tasks. Analysis of the experimental MEG datasets showed that beamformers from the ReciPSIICOS family, but not classical LCMV BFs, detected the expected bilateral focal sources in the primary auditory cortex and detected motor cortex activity associated with audio-motor task, see figure 6. In most cases, MNE performed well, but, as expected, resulted in an excessively smooth solution leading to low spatial resolution. Estimates of the activity of the cerebral cortex using the new BF have several times higher signal-to-noise ratio compared to those based on the classical BFs. This effect is expected and indirectly characterizes the extent to which correlation of sources attenuates the output power as described by expressions 6 and 7.

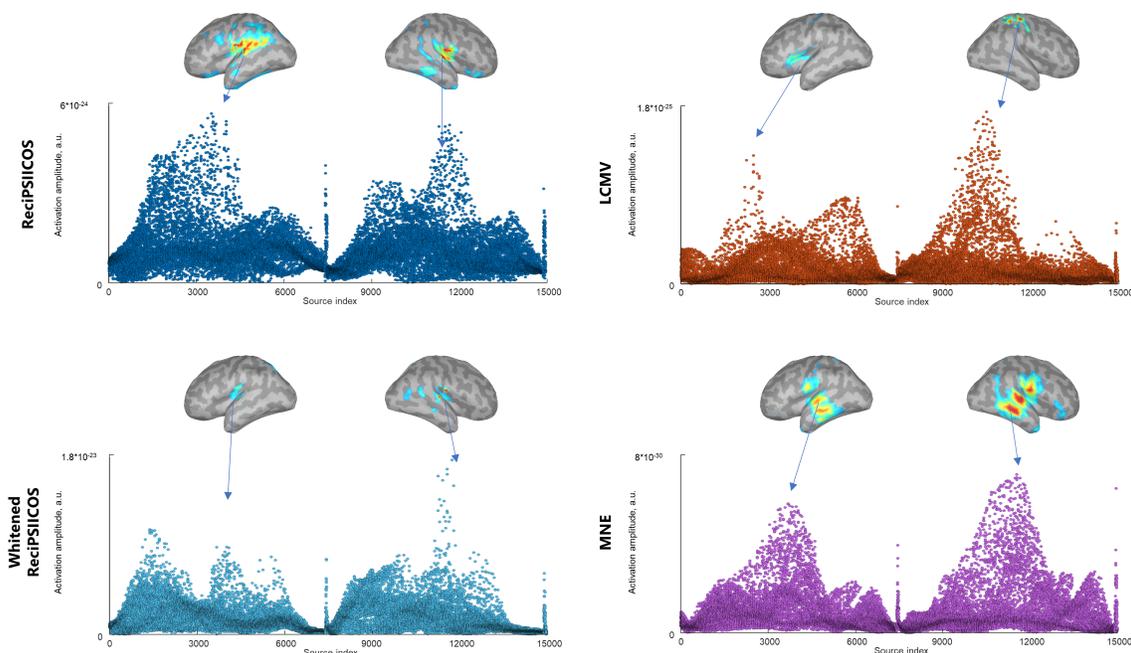


Рис. 6: ASSR power distribution 250 ms post stimulus reconstructed with ReciPSIICOS, Whitened ReciPSIICOS, LCMV and MNE for subject 2

## 3.2 Detection of functional networks with low phase delay from non-invasive measurements of brain activity

### 3.2.1 Introduction

A growing body of evidence suggests that communication between neural ensembles is a defining feature of functionally specialized brain networks and that such communication can be tracked by observing the synchronization between activity of individual brain regions Varela et al. (2001); Fries (2015). Detection of functional networks from electroencephalography (EEG) and magnetoencephalography (MEG) data using

conventional measures of time series synchronicity (such as coherence or phase locking value) is complicated by the presence of volume conductance or spatial leakage when it comes to magnetic field measurements using MEG. Methods such as imaginary coherence [Nolte et al. \(2004\)](#), phase lag index [Stam et al. \(2007\)](#); [Vinck et al. \(2011\)](#), or orthogonalized amplitude correlations [O’Neill et al. \(2015\)](#) eliminate spatial leakage while ignoring the interactions of sources whose activity profiles have zero or close to zero phase difference. While useful, these metrics by design produce false negatives in cases where there is a true zero-phase relationship in the data, and tend to underestimate synchronized sources in situations where the phase delay is close to zero. Ironically, in many cases empirically observed neuronal synchrony is characterized by a vanishing or very small phase delay between the time series of functionally related ensembles of neurons ([Roelfsema et al. \(1997\)](#); [Singer \(1999\)](#); [Engel A.K. \(2001\)](#) ). The true physiological zero latency and such functional connectivity of ensembles with near zero latency can have several explanations. Macroscale analysis shows that two cortical regions involved in bidirectional (reciprocal) interactions are likely to generate near-zero phase lag synchrony as a result of mutually symmetrical interactions [Singer \(1999\)](#); [Engel A.K. \(2001\)](#). An alternative scenario is the presence of a third source (e.g. thalamus) driving both populations [Rajagovindan and Ding \(2008\)](#). Moreover, the near-zero lag could also be due to the effect of dynamical adjustment of the frequency of two populations’ activity allowing them to tune to the global rhythm [Pikovsky et al. \(2001\)](#); [Schuster and Wagner \(1989\)](#).

The new method proposed here allows us to mitigate the undesirable effects of spatial leakage and for the first time detect zero and near-zero phase interactions from non-invasive data. To this end, we generalize the projection operation in the  $M^2$ -dimensional space of sensor signal covariance matrices considered in the previous chapter and project the vectorized cross-spectrum matrix of sensor signals orthogonally to the subspace  $\mathcal{S}_{pwr}^K$ , which makes it possible to suppress the contribution of spatial leakage and, at the same time, partially retain the true component of the interaction with the zero phase. Next, we solve the network estimation task as a multidimensional regression problem in the space of 2-topographies, similar to the traditional source search problem, only in this case, the formal sources are elementary networks with two nodes, and not the neuronal sources themselves, i.e. not separate nodes of these networks, but pairs their pairs whose coherence replaces the role of source activation in the conventional source estimation task. This approach makes it possible to achieve invariance in the phase delay of the time series of functionally connected sources and to ensure a uniform probability of network detection over the entire range of phase difference. Accordingly, our approach is called PSIICOS (Phase shift invariant imaging of coherent sources). Realistic modeling shows that PSIICOS has better detector performance than a number of existing techniques for detecting functional relationships from non-invasive electrophysiological measurements. We also illustrate the performance of PSIICOS applied to a real MEG dataset recorded during a standard mental rotation task. Overall, the proposed approach is a novel method for detecting functional networks from MEG/EEG that overcomes previous limitations and provides invariance to phase delay. Further, the main provisions of the new technique are presented and some results are described. More details about the approach can be found in [Ossadtchi et al. \(2018\)](#).

### 3.2.2 Generating equation of cross-spectral matrix of sensor signals

The complex-valued time-dependent cross-spectral matrix  $\mathbf{C}^{XX}(t, f)$  of sensor signals corresponding to activity at frequency  $f$  is defined as the correlation matrix of the data vector  $\mathbf{x}(t)$  transformed into time-frequency domain as  $\mathbf{X}(t, f) = TFFT(\mathbf{x}(t))$ . The spatial correlation matrix is calculated for each pair  $(t, f)$  and can be formally expressed as

$$\mathbf{C}^{XX}(t, f) = E\{\mathbf{X}(t, f)\mathbf{X}^H(t, f)\} = \Re(\mathbf{C}^{XX}(t, f)) + i \cdot \Im(\mathbf{C}^{XX}(t, f)) \quad (10)$$

By substituting the generating data model  $\mathbf{x}(t)$  into the definition of the cross-spectral matrix (10), representing the matrix products as the sum of the outer products of the columns of the corresponding matrices and separating the imaginary and real parts, similar to the way it was done in (Ewald et al. (2014)) we get:

$$\begin{aligned} \mathbf{C}^{XX}(t, f) = & \boxed{\sum_{i=1}^L \mathbf{g}_i \mathbf{g}_i^T c_{ii}^{ss}(t, f)} + \sum_{i=1}^L \sum_{j=i+1}^L (\mathbf{g}_i \mathbf{g}_j^T + \mathbf{g}_j \mathbf{g}_i^T) \Re(c_{ij}^{ss}(t, f)) + \\ & + i \cdot \left[ \sum_{i=1}^L \sum_{j=i}^L (\mathbf{g}_i \mathbf{g}_j^T - \mathbf{g}_j \mathbf{g}_i^T) \Im(c_{ij}^{ss}(t, f)) \right] + \mathbf{C}^{NN}(t, f) \end{aligned} \quad (11)$$

Here  $\mathbf{C}^{NN}(t, f) = E\{\mathbf{N}(t, f)\mathbf{N}^H(t, f)\}$  is the cross-spectral matrix additive noise of observation, including the contribution from brain activity that is not related to the cognitive process under study and, accordingly, does not contain the components locked to the beginning of a cognitive or motor task under study.  $c_{ij}^{ss}$  - elements of the cross-spectral source matrix  $\mathbf{C}^{ss}(t, f) = E\{\mathbf{S}(t, f)\mathbf{S}^H(t, f)\}$ , whose evaluation is our main goal.

Note that the first term in (11) depends only on the power of sources  $c_{ii}^{ss}$  and  $c_{ii}^{zz}$  and represents the effect of spatial leakage (SL). Since  $c_{ii}^{ss}$  and  $c_{ii}^{zz}$  are real numbers, and  $\mathbf{g}_i \mathbf{g}_j^T - \mathbf{g}_j \mathbf{g}_i^T$  vanishes for  $i = j$  when summed in the imaginary part, the entire SL effect is present only in the real part of the cross-spectral matrix. It is this very observation (Nolte et al. (2004)) that has led to the creation of a variety of methods for non-invasive mapping of functional relationships based on EEG and MEG measurements using only the imaginary part of the cross spectrum, statistics free from the effect of spatial leakage.

Two problems remain unresolved with current methods. First, the terms in the imaginary part of the cross spectrum are modulated by the imaginary part of the original space  $\Im(c_{ij}(t, f))$ . By design, the imaginary part has maximum sensitivity to the 90-degree phase difference of the associated time series. Consequently, pairs of coherent sources with an average phase difference close to zero turn out to be practically invisible in the imaginary part of the cross spectrum, especially taking into account the presence of additional natural noise in it. Second, while the real part of the cross-spectrum of the sensory space will have the maximum contribution from any true zero-phase interactions, it is also contaminated by the effect of spatial leakage, which negatively affects the resulting signal-to-noise ratio.

### 3.2.3 New method

It is easy to see that this problem is complementary to the one that we solved in the previous section, in which our task was precisely to suppress the contribution to the covariance matrix of the cross terms modulated by the degree of source correlation and to emphasize the contribution from the components modulated by the power of the sources. Now our task is exactly the opposite and we approach it also using the projection in  $M^2$  - the dimensional space-product of the sensor signals. However, we project not onto  $\mathcal{S}_{pwr}^K$ , but away from it, in order to suppress the contribution of spatial leakage into the cross-spectrum matrix.

Thus, to form the projection matrix, we use a procedure similar to that described in the previous section and illustrated in Figure 3, calculate the matrix  $\mathbf{P}$  and its complementary matrix  $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P}$ , which is then applied to the vectorized cross-spectral matrix. In general, the algorithm can be described as follows:

- We start with the forward model matrix  $\mathbf{G} = \{\mathbf{g}_{ij}\}$ ,  $i = 1, \dots, K$ ,  $j = 1, \dots, N$  corresponding to all  $N$  nodes of the cortical grid and form the  $M^2 \times N$  matrix  $\mathbf{G}_{pwr}$  from auto 2-topographies  $\mathbf{q}_{ii} = \text{vec}(\mathbf{g}_i \mathbf{g}_i^T)$ ,  $i = 1, \dots, N$ .

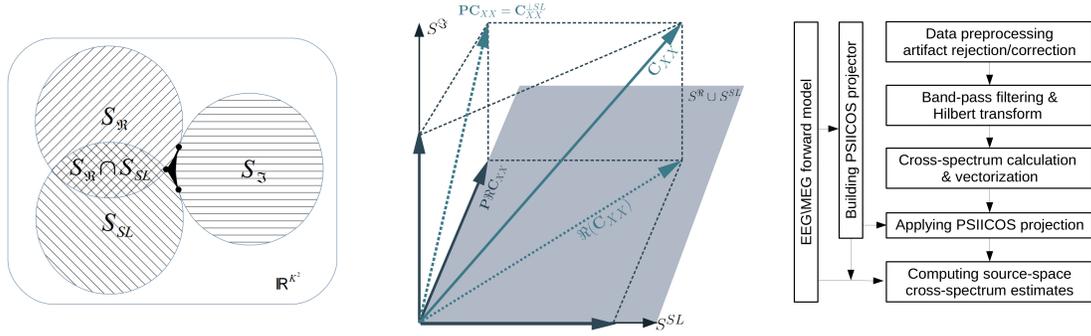
$$\mathbf{G}_{pwr} = [\mathbf{q}_{11}, \mathbf{q}_{22}, \dots, \mathbf{q}_{NN}] \quad (12)$$

- We form the projection matrix from the subspace  $\mathcal{S}_{pwr}^K$  as  $\mathbf{P}^\perp = \mathbf{I} - \mathbf{U}_R \mathbf{U}_R^T$ , where  $\mathbf{U}_R = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_R]$  - the matrix of the first  $R$  left singular vectors  $\mathbf{G}_{pwr}$ , defining the  $R$ -dimensional subspace  $\mathcal{S}_{pwr}^K$ , in which the main variance of the spatial leakage is concentrated. This follows from the defining property of SVD. The procedure and criteria for choosing the rank  $R$  of the projection operator are described in detail in [Ossadtschi et al. \(2018\)](#).
- We use the projection operator  $\mathbf{P}^\perp$  to suppress the contribution of the leakage effect to the cross spectrum of the sensor space  $\text{vec}(\mathbf{C}^{XX})(t, f)$ :

$$\text{vec}(\mathbf{C}^{XX})^{\perp SL}(t, f) = \mathbf{P}^\perp \text{vec}(\mathbf{C}^{XX})(t, f) \quad (13)$$

The 14 equation resembles the equation for observing sensor signals in the first part, however, in the current case, the observed quantity is the cross spectrum vectorized and projected from the volume conduction subspace, and instead of actual sources we have elementary 2-networks connecting two neuronal populations in the cerebral cortex and their coherence  $c_{ij}^{ss}(t, f)$  that is to be estimated.

This problem can be solved by various approaches, from simple scanning using the methods of the MUSIC (Multiple signals classification) [Mosher and Leahy \(1999b\)](#) family to solving a global optimization problem taking into account the structural connectome model and imposing a smooth profile of variability in time on the degree of synchronization of a particular network, which can be achieved using mixed rules. Here, however, we restrict ourselves to a simple MUSIC method by calculating the cosine of the angle between the projected 2-topographies  $\mathbf{q}_{ij}^\perp$  and the projected and vectorized slice of the cross-spectrum matrix for each time  $t$  and for a fixed frequency  $f$   $\text{vec}(\mathbf{C}^\perp(t, f))$ :



(a) Venn diagram of the relationship of subspaces into which the cross-spectral matrix can be decomposed.

Thus, by suppressing the contribution of the spatial leakage component (or volume conductance in the case of EEG), we have essentially obtained a regression problem for estimating  $c_{ij}^{ss}(t, f)$  from the data of the projected cross-spectrum  $vec(\mathbf{C}^\perp)(t, f)$  based on the following observation equation

$$vec(\mathbf{C}^\perp)(t, f) = \sum_{i=1}^L \sum_{i \neq j, j=1}^L \mathbf{q}_{ij}^\perp c_{ij}^{ss}(t, f) + \epsilon(t, f) \quad (14)$$

where  $\mathbf{q}_{ij}^\perp = \mathbf{P}\mathbf{q}_{ij}$  are projections of 2-topographies, and  $\epsilon(t, f)$  additive noise including residuals of spatial leakage.

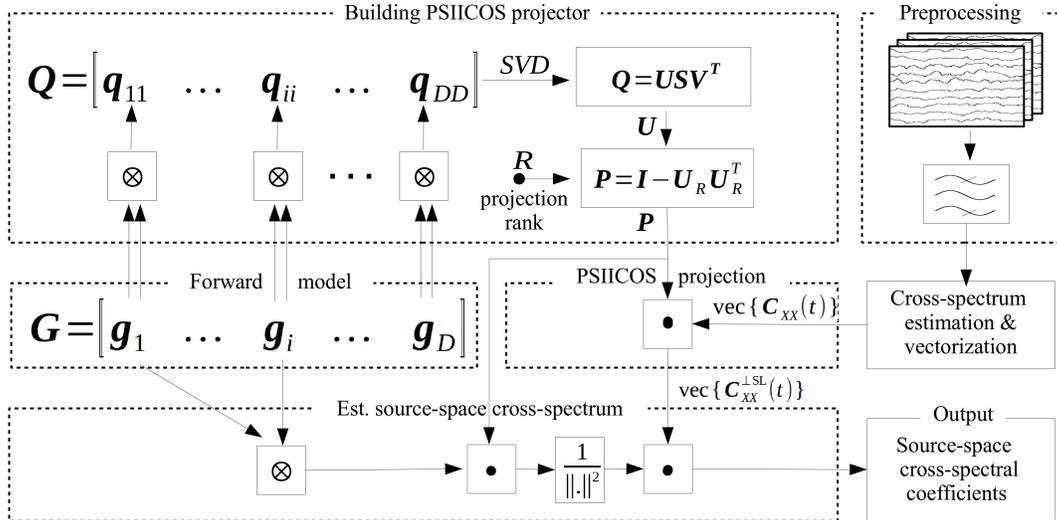


Рис. 7: Graphical representation of the PSIICOS technique

$$\rho_{ij}(t, f) = \langle \mathbf{q}_{ij}^\perp / \|\mathbf{q}_{ij}^\perp\|, \text{vec}(\mathbf{C}^\perp(t, f)) / \|\text{vec}(\mathbf{C}^\perp(t, f))\| \rangle \quad (15)$$

Inspired by the ideas of the recursively applied RAP-MUSIC [Mosher and Leahy \(1999a\)](#) here, after doing a full scan over  $N(N - 1)$  elementary 2-nets, we choose a pair  $(i_0, j_0)$  with a maximum value of  $\rho_{ij}(t, f)$ , so that  $(i_0, j_0) = \text{argmax} \rho_{ij}(t, f)$  and project our data and 2-topographies model matrix into the space orthogonal to  $\mathbf{q}_{i_0 j_0}^{\text{perp}}$  and repeat the scanning procedure in search for the global maximum of  $\rho_{ij}(t, f)$ . The iterations stop after the observed value  $\max \rho_{ij}(t, f)$  falls below the threshold  $\theta = 0.7$ , whose value we derived empirically.

### 3.2.4 Results

Figure 8 shows the result of applying the projection operator to 2-topographies that contribute to the real and imaginary components of the cross spectrum corresponding to the interaction of sources,  $\mathbf{q}_{ij} + \mathbf{q}_{ji}$  and  $\mathbf{q}_{ij} - \mathbf{q}_{ji}$  respectively, as well as to the component due to spatial leakage. We are primarily interested in weakening the contribution of the spatial leakage components and the maximum preservation of the cross-spectrum components modulated by the imaginary and real parts of the off-diagonal elements of the cross-spectral matrix of sources. Figures 8.a) and c) show the norms of the original (before projection) 2-topographies corresponding to the three listed components (see also 11), for all possible source pairs  $(i, j)$  and located depending on the angle between the topographies  $\mathbf{g}_i$  and  $\mathbf{g}_j$  - panel a), and depending on the distance between the coordinates of the  $i$ -th and  $j$ -th nodes of a discrete cerebral cortex grid. Similar plots, but for projected 2-topographies, are shown in panels b) and d).

As can be seen from the diagrams, the component responsible for the volume conductivity (the cloud of yellow dots) is weakened by the projection by more than a factor of 10 and is practically pressed against the horizontal axis. In this case, the cloud of norms of 2-topographies (blue color), modulated by the real part of the cross-spectral coefficients of the sources, changes its shape and the ordinates of the points are shifted in the direction of decreasing the norm. However, most of the points do not experience more than 2-fold cancellation, which on average corresponds to greater than 5-fold increase in the signal-to-noise ratio for networks with zero phase in the projected cross spectrum compared to the original cross spectrum. In this case, of course, those networks whose nodes are close to each other and whose node topographies are highly correlated undergo maximum attenuation, since their 2-topographies are most similar to the auto 2-topographies corresponding to spatial leakage.

To study the properties of the new method and compare it with several existing approaches, we conducted a series of numerical experiments using realistic simulations. The simulation details and results are presented in [Ossadtchi et al. \(2018\)](#), here we describe only a few key results.

A Monte Carlo study was conducted in which a single network with two nodes was generated, active for half of the time window under study. The network nodes at each iteration were in random positions. Two scenarios of phase connectivity were modeled. In the first, the phase difference of the oscillations was equal to  $0.2\pi$ , and in the second -  $\pi/2 - 0.2\pi$ . The task was to detect the network and find its nodes from synthetic MEG data, which was formed from the activity of the network mixed with additive realistic spatially colored noise generated by the activity of 500 neuronal populations.

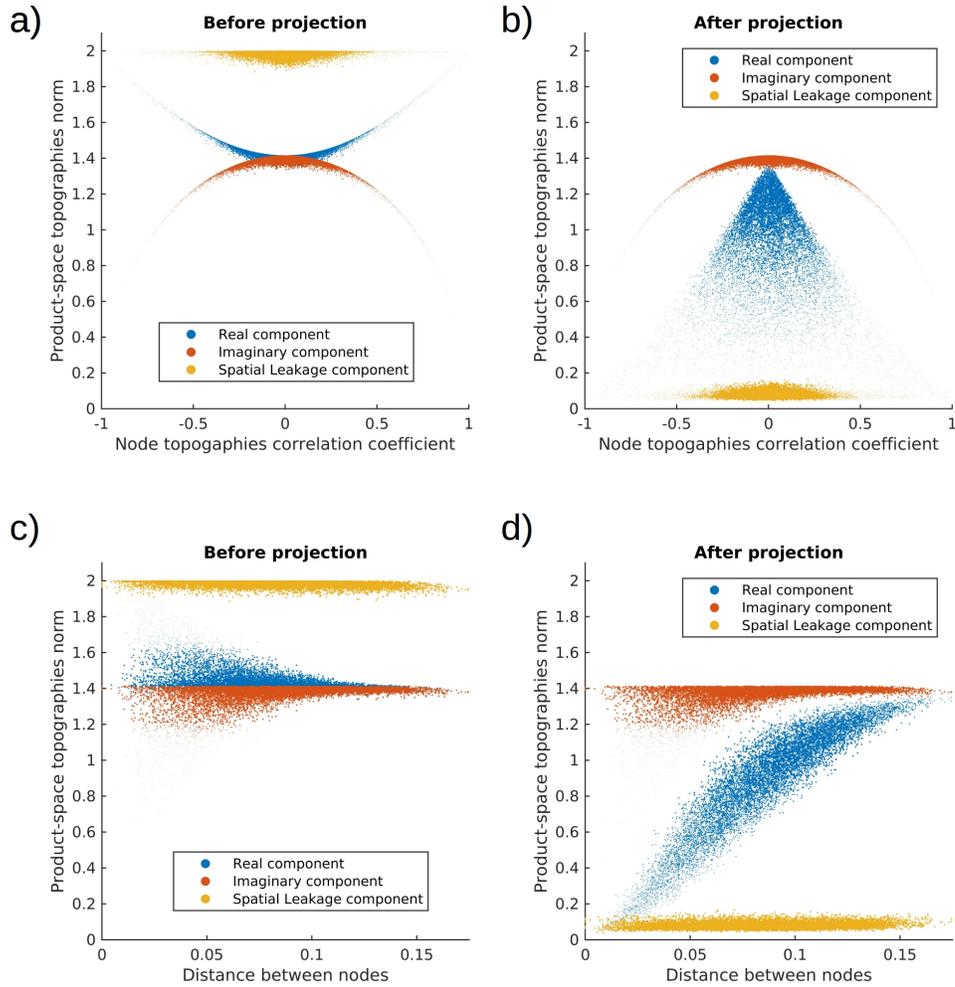


Рис. 8: Product space topography norms for three subspaces of the cross-spectrum of sensor signals before and after the PSIICOS projection depending on the correlation coefficient of the initial topographies of connected nodes (panels a) and b)) and the distance between a pair of elementary network nodes (panels c) and d)). Before projection (panels a), c), the source power component (yellow) dominates in the cross-spectrum of the sensory space. After the projection (panels b, d)) the manifestation of the source power on the sensors is reduced by at least a factor of 10. We also observe an inevitable but much less abrupt decay of the average norm of the topographies corresponding to the real part of the cross spectrum.

The proposed approach was compared with three other methods most commonly used to analyze functional relationships in EEG and MEG data, DICS, iDICS and MNE GCS. The performance of the methods was evaluated using Precision-Recall and ROC curves. The network was considered correctly detected if the distance from each of its true nodes to the corresponding network node detected by the algorithm did not exceed 1 cm.

As can be seen from Figure 9 PSIICOS demonstrates significantly higher performance compared to the three existing approaches in a wide range of signal-to-noise ratios, both at low and maximum phase delays.

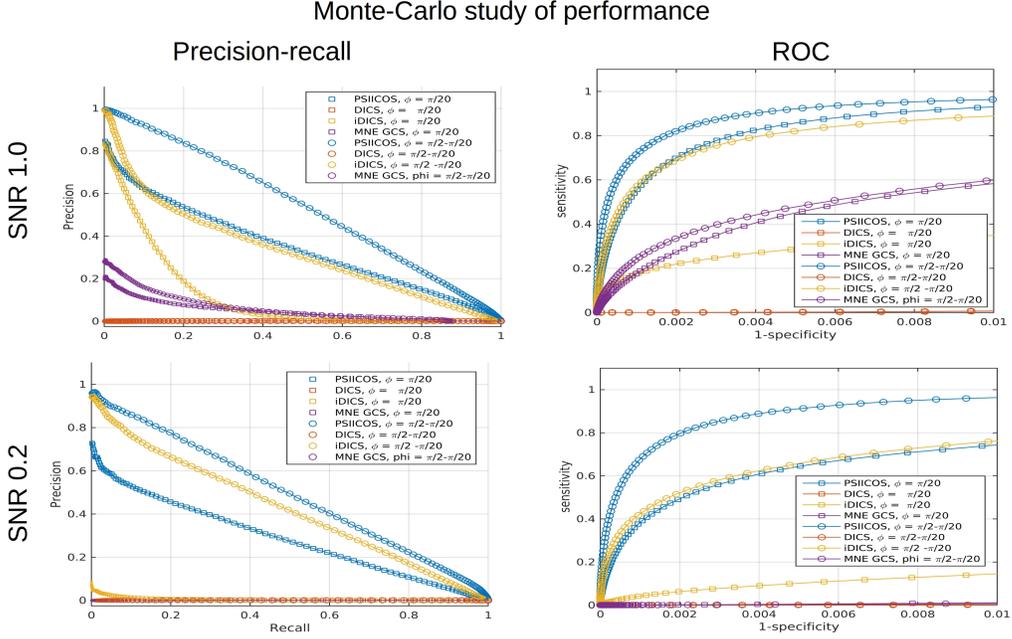


Рис. 9: PR curves (a,c) and ROC (b,d) curves comparing the performance of network detectors; PSIICOS - DICS, iDICS and GCS MNE approaches for two different noise levels.

Next, we modeled three networks with nodes located in different hemispheres. These networks were activated in three different overlapping windows of the analyzed time interval, as shown in Figure 10. As before, we considered two phase angles of delay and two levels of noise in the data. The results of 4 algorithms for 4 conditions are shown in Figure 11. As can be seen from the graphs, PSIICOS reliably detects all three networks under all simulated conditions. It is also possible to estimate the time windows in which each of the networks is active, see Figure 12.

Cleaning real part of the cross-spectrum from the contribution of spatial leakage using the proposed projection operation makes it possible to ensure the invariance of network detector's performance to the the phase lag between the activity of functionally related populations. Figure 13 shows the result of the simulation of the detector for networks with different phase delays. The graphs show three curves corresponding to a detector operating only on the real component projected from the spatial leak, only on the imaginary one, and on the projected complex cross spectrum. As expected, the detector of networks based on the imaginary part of the cross spectrum has poor detector performance for small phase angles. The reverse situation is observed when using the real component of the cross-spectrum, cleared of spatial leakage using the PSIICOS projection. The use of both the imaginary and the cleaned real component makes it possible to significantly reduce the dependence of performance on the phase

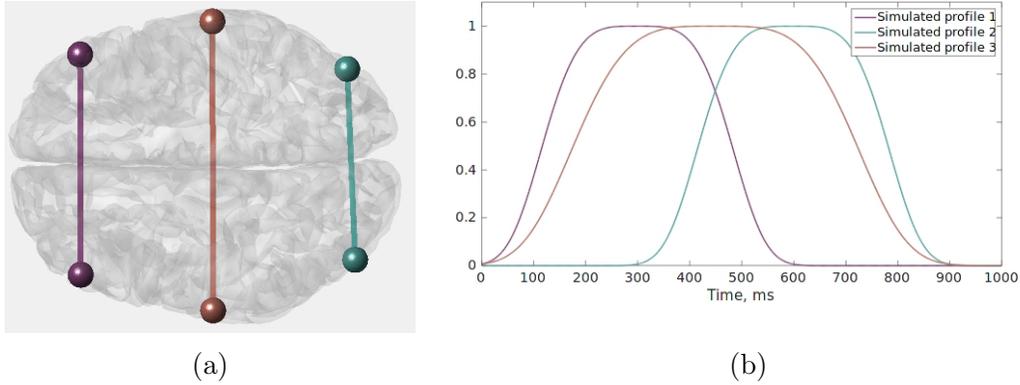


Рис. 10: Test with three pairs of synchronous sources a) - spatial structure of networks, b) - temporal activity profiles of three networks.

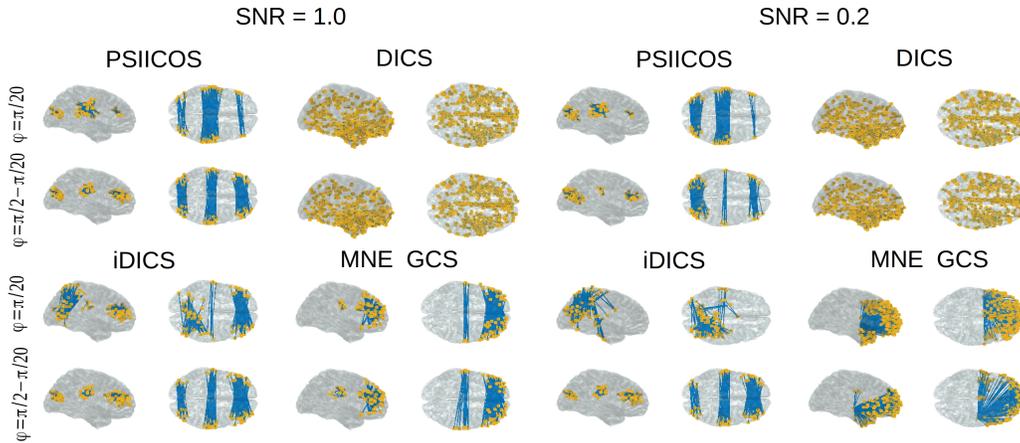


Рис. 11: Spatial structure of networks reconstructed using 4 methods for two different values of phase shift and noise level.

angle of synchronization.

Finally, we applied PSIICOS to the analysis of real MEG data recorded from a subject involved in a mental rotation task. As part of the task, the subject was randomly presented with an image of the right or left hand, rotated randomly. The subject's task was to decide which hand is shown in the presented picture. To do this, it is necessary to perform the operation of mental rotation of the imagination and compare the result of such rotation with a more canonical internal image of the hand. Using the bootstrap procedure, stable networks were identified in the alpha, beta and low and high gamma ranges, see Figure 14, the spatial and temporal characteristics of some of them are shown in Figure 15 .

The physiology of the observed networks, with a dominant contribution to the real component of the cross-spectrum, leaves no doubt. Network 1 corresponds to the interaction between sections of the ventral visual pathway and motor areas of the cerebral cortex, present in the traditionally motor beta range. Network 2 represents

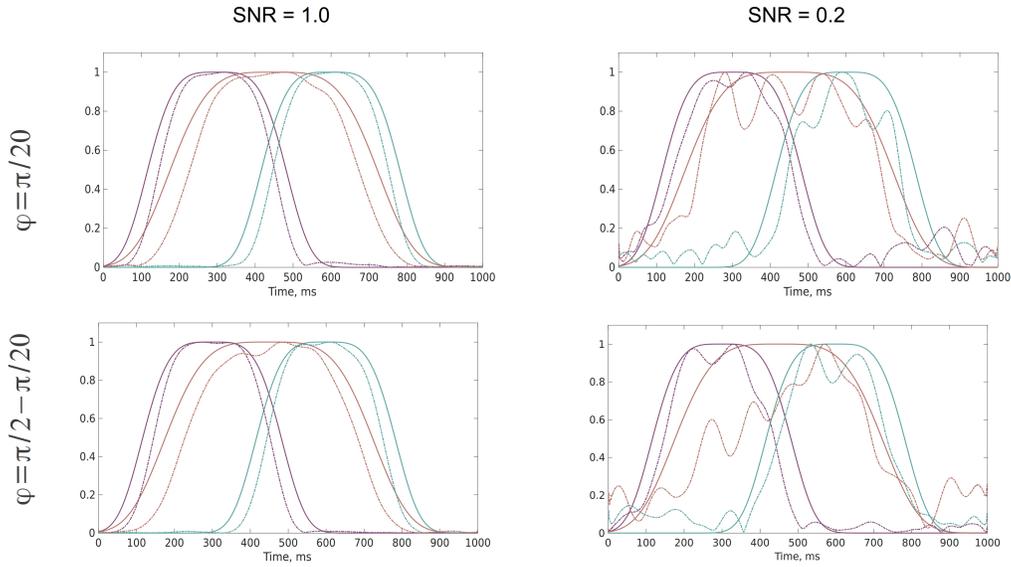


Рис. 12: Temporal activity profiles of each of the three networks, estimated using PSIICOS, superimposed on the true intensity profiles of communication within each of their networks.

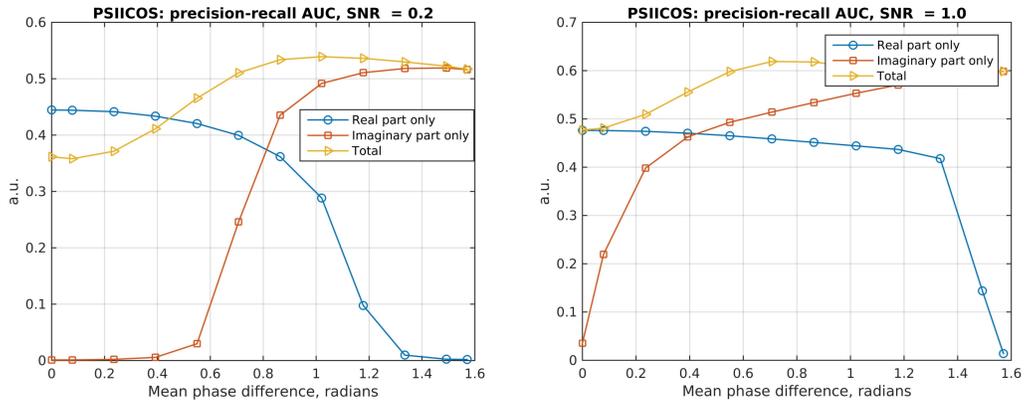


Рис. 13: Network detector performance

the exchange of information between two homologous sensorimotor regions in the two hemispheres.

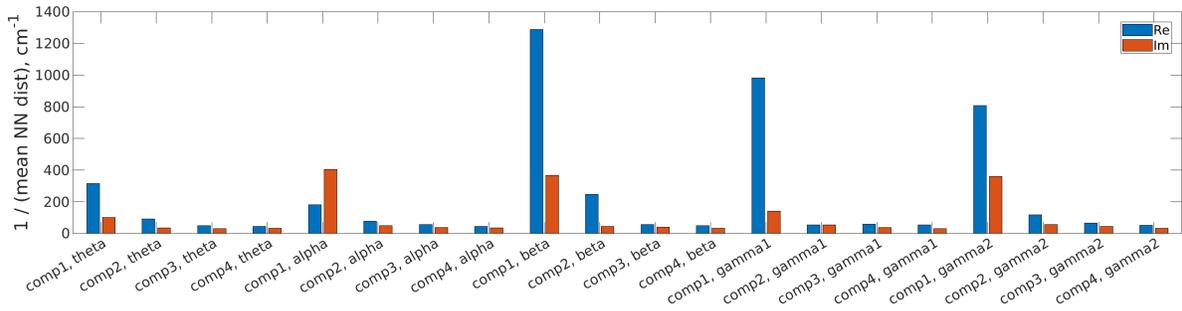


Рис. 14: Results of bootstrap analysis of network reproducibility. To assess the reproducibility of the MUSIC scan results, a multiple selection of a subset of data epochs was carried out, followed by the calculation of the cross-spectrum tensor by averaging over these epochs, to which the PSIICOS projection procedure was then applied and then a RAP-MUSIC scan was performed. The height of the bar corresponds to the reproducibility index in a particular frequency range, calculated as the reciprocal of the average distance to the nearest neighbor for networks found in the real (blue) and imaginary (red) parts of the cross spectrum. Accordingly, the higher the bar, the closer and more similar were the networks found at each of the bootstrap iterations.

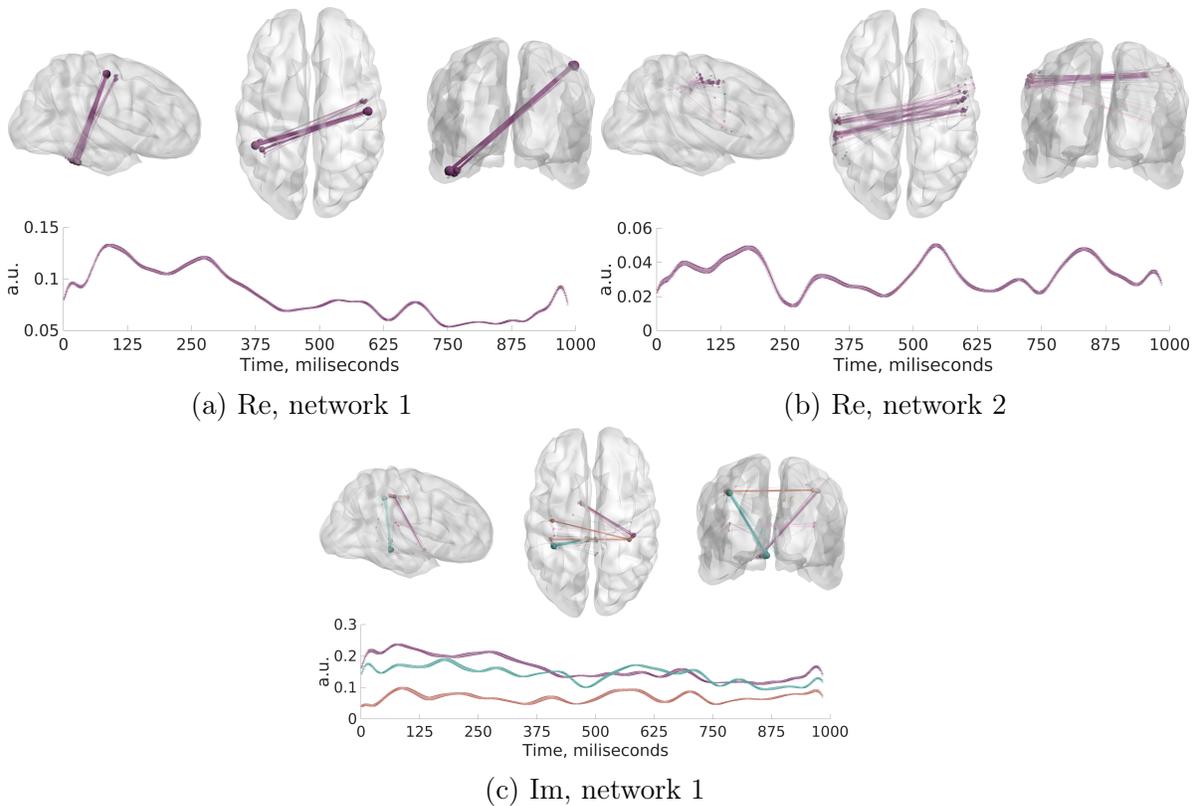


Рис. 15: Spatial and temporal dynamics of the most reproducible networks observed in the beta (16-24 Hz) range.

## 4 New processing methods applied to the diagnosis of epilepsy

### 4.1 Efficient biomimetic method for detecting interictal discharges (spikes) in multichannel recordings of brain electrical activity.

According to the World Health Organization (WHO), epilepsy is one of the most common neurological disorders. In combination with such clinical manifestations as seizures, this disorder is characterized by specific bursts of EEG/MEG activity - the so-called interconvulsive spikes, due to the synchronous activity of groups of neurons in epileptogenic areas of the brain (Staley and Dudek, 2006). Interictal adhesions (this is a medical term accepted in the Russian Federation) are considered a reliable biomarker of epilepsy, which can be detected non-invasively and used to localize the irritative zone. The development of automatic spike detection algorithms is in demand due to the fact that traditional manual analysis is accompanied by such disadvantages as high time costs and low consistency of spike detection results between experts (Webber et al., 1993; Hostetler et al., 1992; Halford et al., 2013; Scheuer et al., 2017).

Existing approaches to automatic spike detection include mimetic methods based on the decomposition of the spike shape into constituent elements (Gotman and Gloor, 1976; De Oliveira et al., 1983; Faure, 1985; Glover et al., 1986; Gotman and Wang, 1991, 1992; Dingle et al., 1993; Keshri et al., 2011; Liu et al., 2013); methods using mathematical morphology Nishida et al. (1999); pattern matching, as well as methods based on wavelet decomposition or decomposition into independent components Ossadtschi et al. (2004a). We have recently introduced the approach dubbed fast parametric curve matching (FPCM (Kleeva et al., 2022)) based on a constrained mixed spline model and characterized by mimetic principles and scale independence. Within this method, the spike shape  $s(t)$  is determined by two linear segments and parabola, see figure 16 :

$$s(t) = \begin{cases} c_1 t + c_2 & \text{if } -N_1 \leq t \leq 0, \\ c_3 t + c_4 & \text{if } 0 \leq t \leq N_2, \\ c_5 t^2 + c_6 & \text{if } N_2 \leq t \leq N_2 + N_3, \end{cases} \quad (16)$$

(17)

where  $c_1$  and  $c_3$  correspond to the slope of the linear segments of the spike peak,  $c_2$  and  $c_4$  correspond to segments,  $c_5$  scales the parabolic curve approximating the wave, and  $c_6$  is the corresponding segment.

For fitting morphological model  $\mathbf{B}$ , the residual square error is minimized. The coefficients of the polynomial spline correspond to the solution of the following equation:

$$\mathbf{c}(t) = \operatorname{argmin} (\|\mathbf{B}\mathbf{c} - \mathbf{x}(t)\|^2), \quad (18)$$

where  $\mathbf{x}(t)$  is the EEG/MEG data segment around the spike peak,  $\mathbf{c}(t)$  is the spline coefficient vector for the specified segment, and  $\mathbf{B}$  is the morphological matrix models. The construction of the morphological model  $\mathbf{B}$  is implemented as follows. As a first

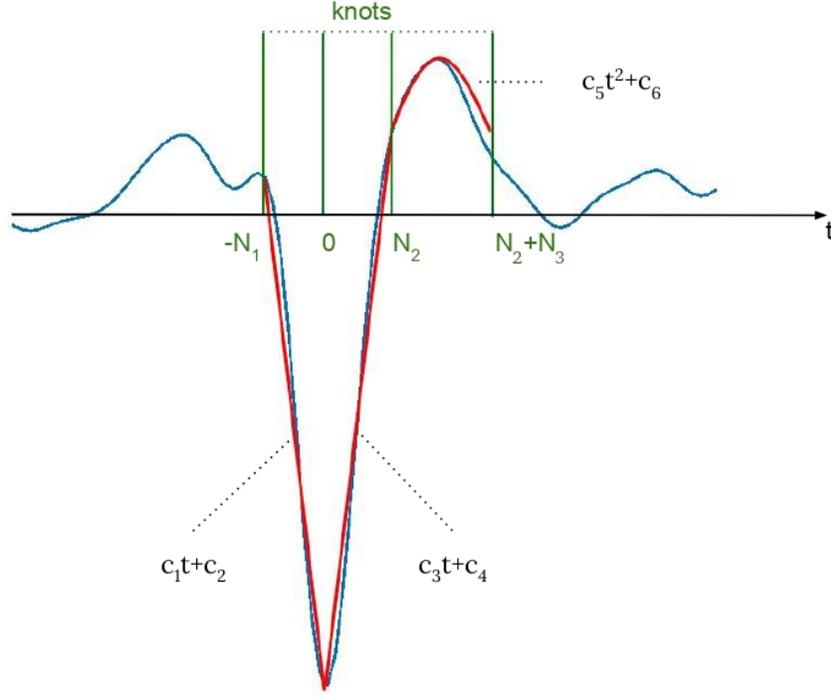


Рис. 16: Spike shape parameterization (blue) using spline (red)

step, three elements of the forward model are defined corresponding to three elements (the first segment of the peak, the second segment of the peak and the wave):

$$\mathbf{A}_1 = \begin{bmatrix} -N_1 & 1 \\ -N_1 + 1 & 1 \\ -N_1 + 2 & 1 \\ \dots & \dots \\ -1 & 1 \end{bmatrix} \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ \dots & \dots \\ N_2 - 1 & 1 \end{bmatrix} \quad \mathbf{A}_3 = \begin{bmatrix} \left(\frac{N_3}{2}\right)^2 & 1 \\ \left(\frac{N_3 - 1}{2}\right)^2 & 1 \\ \left(\frac{N_3 - 2}{2}\right)^2 & 1 \\ \dots & \dots \\ \left(\frac{N_3 - 2}{2}\right)^2 & 1 \\ \left(\frac{N_3 - 1}{2}\right)^2 & 1 \\ \left(\frac{N_3}{2}\right)^2 & 1 \end{bmatrix} \quad (19)$$

where  $\mathbf{A}_1$  and  $\mathbf{A}_2$  correspond to the peak, and  $\mathbf{A}_3$  to the wave. The first column of each matrix represents the abscissa raised to the power of the spline, where  $N_1$ ,  $N_2$  are the initially specified duration of the linear segments in samples, and  $N_3$  is the duration of the slow wave. A spline can be represented as  $\mathbf{x} = \mathbf{B}\mathbf{c}$ , where

$$\mathbf{B} = \begin{bmatrix} \mathbf{A}_1 & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 \\ 0 & 0 & \mathbf{A}_3 \end{bmatrix} \quad (20)$$

To align the ordinates of the last element of the first linear segment with that of the first element of the second linear segment and to match the last element of the

second linear segment with the first element of the wave, the model is supplemented with additional boundary conditions.

The coefficients corresponding to the three parts of the parameterized spike are obtained from the convolution of the indicated morphological model and the time series corresponding to the EEG/MEG data from each channel. The final decision as to whether a candidate profile is a spike is made based on a set of logical predicates. For example, for spikes with negative polarity, it is expected that the peak is directed towards negative values ( $c_1(t) < 0$  and  $c_3(t) > 0$ ); the top of the peak is below zero ( $h_p(t) = c_4(t) < 0$ ); the wave is directed upwards ( $h_w(t) = c_5(t) < 0$ ); the top of the wave is above zero ( $c_6(t) > 0$ ); the peak height is at least  $\alpha$  times greater than the wave height ( $|h_p(t)| > \alpha|c_6(t)|$ ). An additional criterion for selecting spikes is the requirement for the spike being present in multiple channels. An important property of the presented algorithm is its scale independence. The initial determination of the length of the linear segments and the wave makes it possible to detect spikes on various scales.

The method's diagram is shown in 17. Based on the verbal description, a model of mixed splines is created and a set of logical predicates is formulated. After projecting the mixed spline model matrix from the continuity constraint equations introducing continuity constraints, the spline model matrix is inverted using SVD and FIR filters are formed to quickly calculate the spline coefficients for each time point using a highly efficient convolution operation. After that, the obtained spline coefficients are converted back into a vector of the interpretable coefficients  $\mathbf{c}(t)$ , and for each time interval, the residual error is estimated along with logical predicates. Both operations are easily vectorized, which significantly increases the computational efficiency of the method.

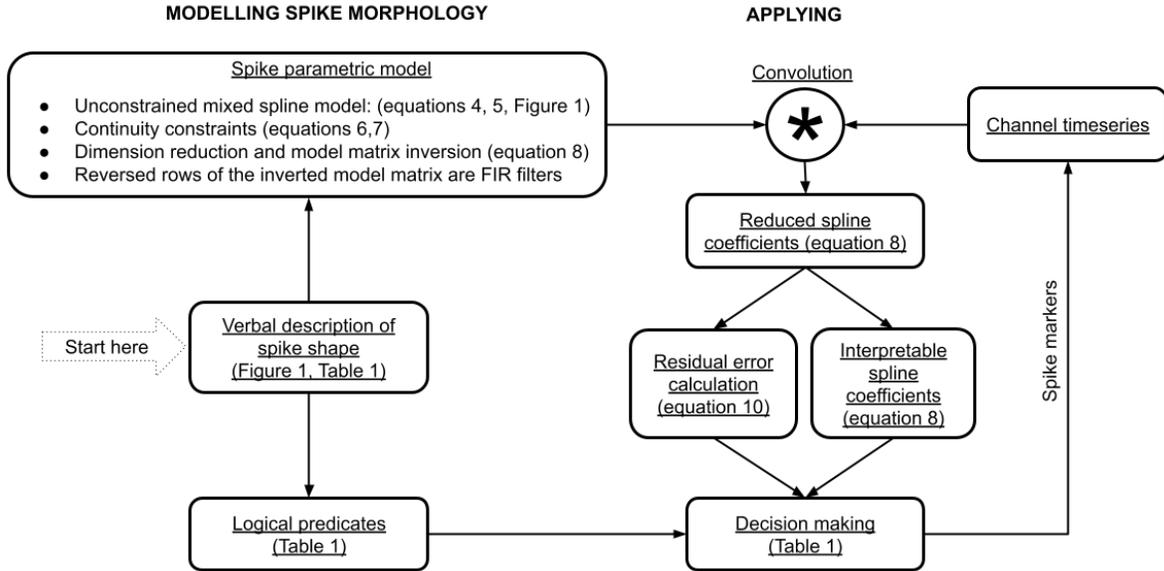


FIG. 17: Data processing pipeline with FPCM.

First, the FPCM algorithm was tested on a realistic simulation dataset in which three brain structures (right anterior transverse temporal gyrus, left parietal cortex, and left parahippocampal gyrus) generated spikes of various shapes. To assess the stability of the algorithm to artifacts, some of the data epochs were accompanied by artifacts emanating from other zones. The amplitude of these artifacts determined the signal-to-

noise ratio (SNR) of the simulated data, see Figure 18.

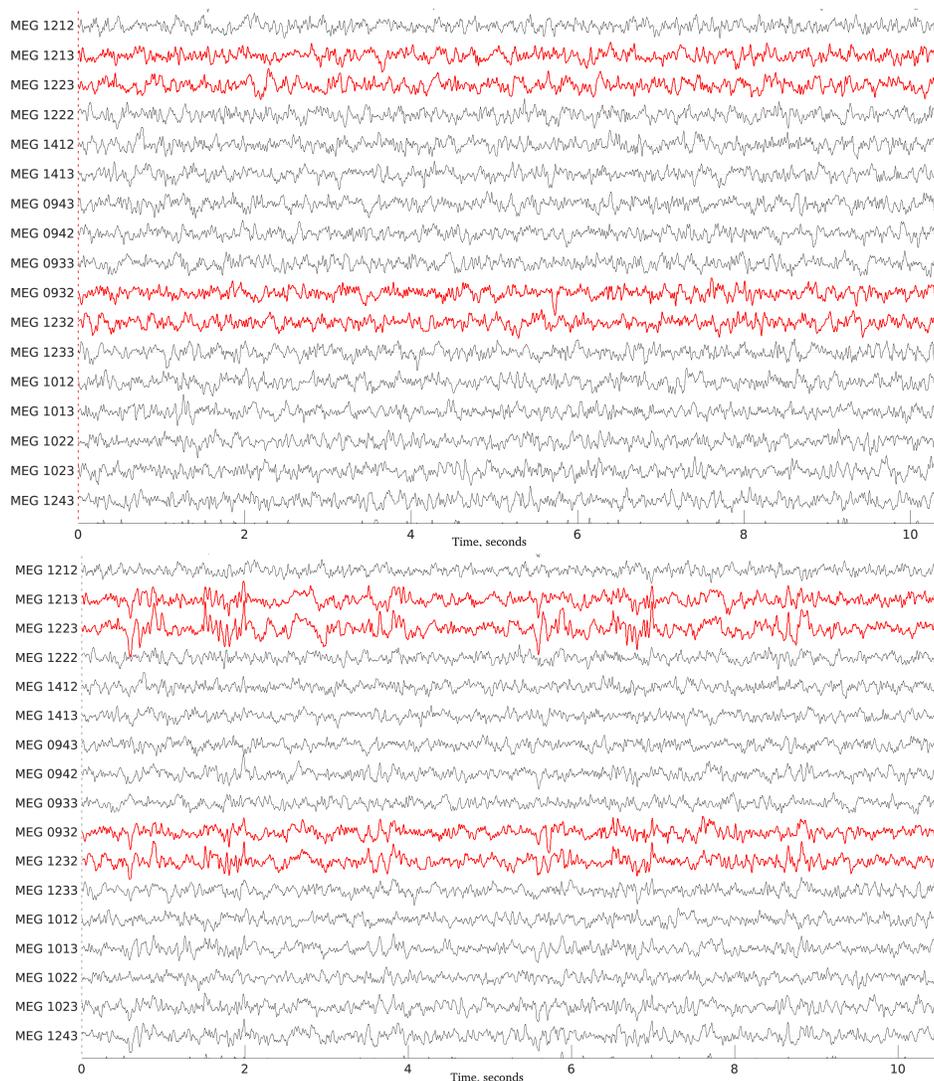


Рис. 18: An example of model data with the minimum and maximum amplitude of artifacts, respectively.

Applying the FPCM algorithm to model data and comparing it with two standard approaches - matched filtering and wavelet decomposition - we observed the superiority of the proposed approach in terms of ROC AUC criteria for all SNRs and resistance to high-amplitude artifacts, see Figure 19. Starting from SNR=1, FPCM outperforms standard approaches, showing the highest AUC (98.2441) versus AUC = 97.9853 for wavelet decomposition and AUC = 97.7368 for matched filtering.

The spatial characteristics of the detected foci of interictal spike generation corresponded to the simulated ones, see Figure 20.

The application of the FPCM algorithm to real data of three patients with drug-resistant epilepsy made it possible to detect spikes corresponding to true irritative zones. As an example, we present here Figure 21, which shows the results of the localization of interictal discharges detected using FPCM.

To demonstrate the robustness of the FPCM method to spontaneous high-amplitude

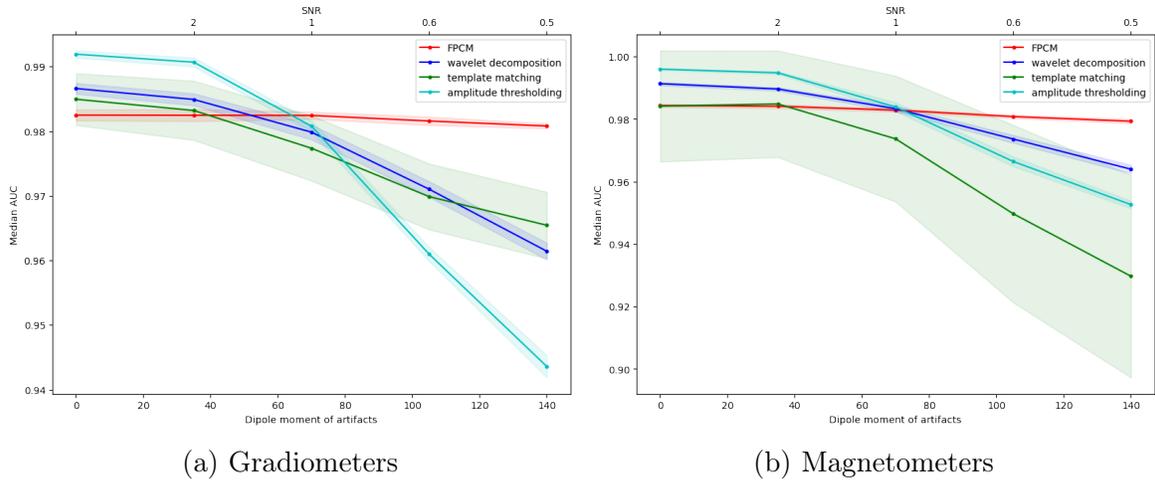


Рис. 19: ROC AUC values versus signal-to-noise ratio (SNR)

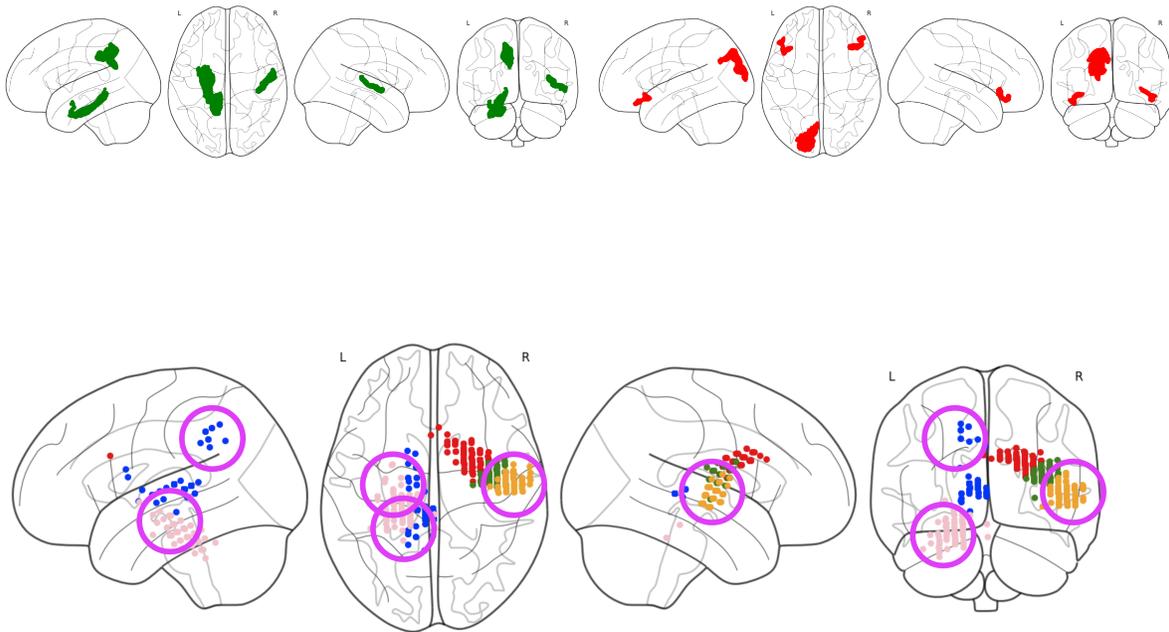


Рис. 20: (a) Zones of simulated spikes. (b) Zones of simulated artifacts. (c) Results of localization of spikes detected using FPCM.

artifacts, we applied FPCM and two other commonly used approaches to EEG data recorded from a patient in the intensive care unit. This short data set included 12 channels with 11 manually detected peaks that occurred at 2 different time points and were recorded on several channels at the same time, see fig. 22.

The proposed FPCM method found 7 true bursts and 2 false bursts. As a result, both true time points were detected with only a single false detection on one channel per 1 minute of data. Importantly, the FPCM does not use a threshold based on the

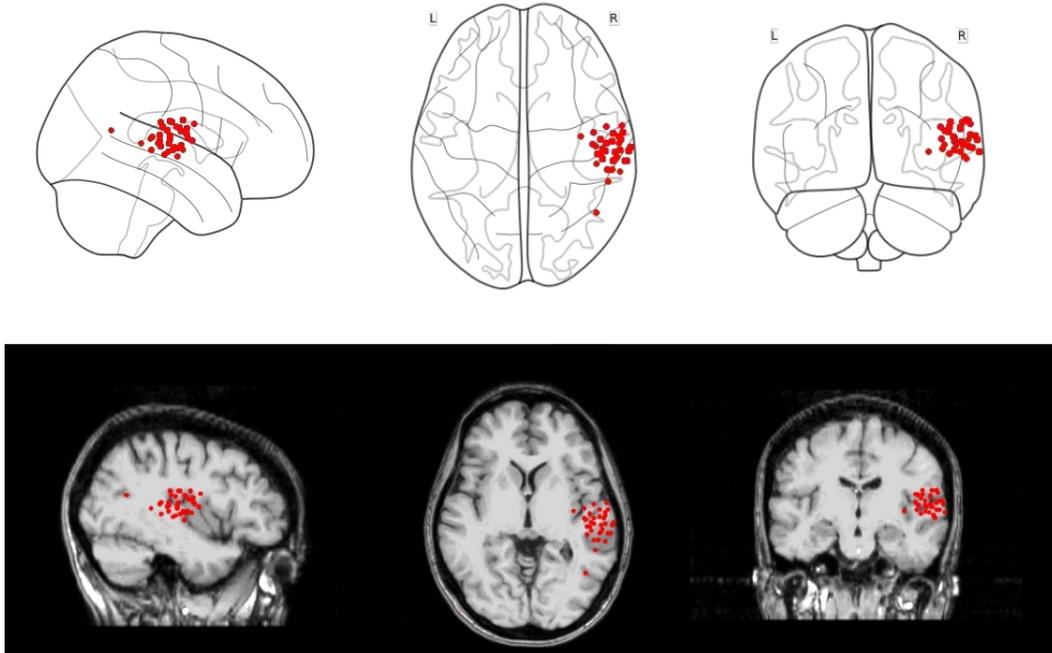


Рис. 21: An example of the localization of the patient's spikes.

absolute magnitude of the peak. Thus, the demonstrated sensitivity is 77.78% if the task is to detect bursts in each channel independently, and 100% if we aim to detect only the times when the desired discharges occur. Note that FPCM also detected one low-amplitude false-positive event, which is due to the fact that FPCM operates solely on the basis of the shape parameter and not the absolute value.

The wavelet transform approach turned out to be too sensitive to high-amplitude artifacts and could not detect any of the true spikes (sensitivity was 0 %). At the same time, 63 false discharges were found. As a result, 21 time points were falsely detected within 1 minute.

For the matched filtering method, we used several patterns based on peaks previously detected using FPCM. Although the matched filtering technique allowed some of the true events (4 true positives) to be detected, 7 items were missed and 41 segments were erroneously labeled as a spike. The resulting sensitivity was 36.36%.

The application of the FPCM to noisy human EEG data and to epileptic activity of the rat's brain recorded invasively demonstrated the superiority of this algorithm over standard approaches, which indicates the ease of generalizing to new data. Thus, since the described algorithm is characterized by low computational complexity, resistant to high-amplitude artifacts, and provides flexibility in describing the spike morphology, it successfully supplements the arsenal of existing tools for analyzing interictal activity.

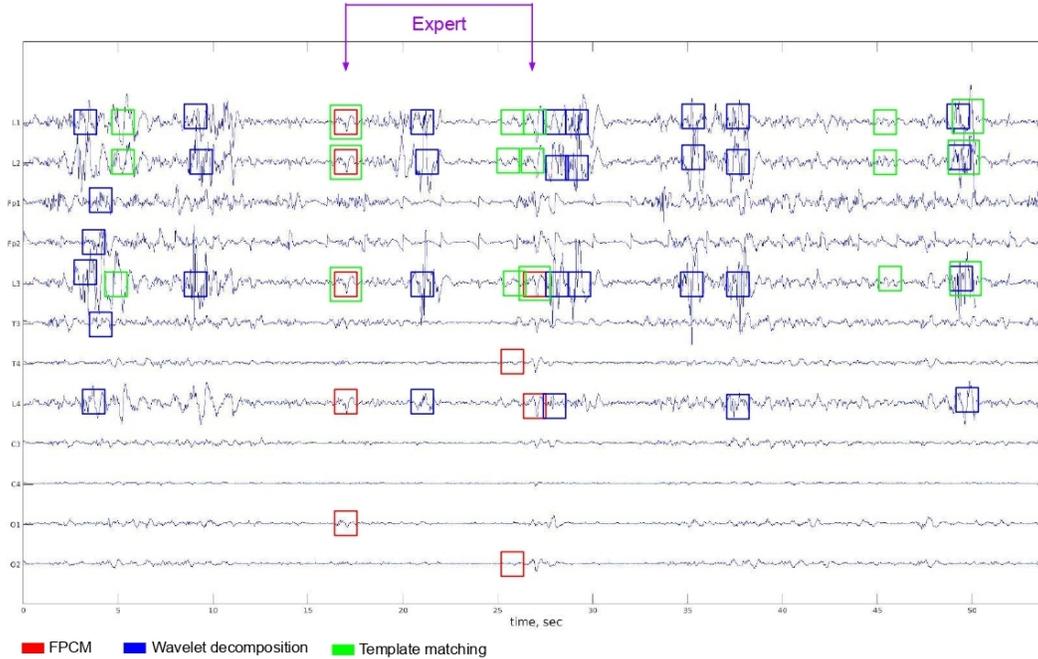


Рис. 22: Comparison of the spikes detected by FPCM, wavelet transform method and matched filtering method. Purple arrows mark interictal discharges (spikes) detected by the expert. The time series also exhibits large high-frequency artifacts that obscure much less visible target events noted by the expert.

## 4.2 Traveling wave model for analysis of local dynamics of interictal discharges propagation and its application to determination of the epileptogenic zone

This section provides a brief summary of a new method for solving the EEG/MEG inverse problem, which is based on the assumption about cortical traveling waves. We explored the properties of the method using realistic simulations. We then applied the proposed algorithm to interictal spikes in the MEG recordings of patients with epilepsy. In addition to being applied to the analysis of interictal discharges, the proposed method can be used to study any neural process that involves wave-like propagation of activity through the cerebral cortex. The method is described in detail in [Kuznetsova and Ossadtchi \(2022\)](#).

### 4.2.1 Data Model

Here we propose a methodology for a non-invasive study of the fine spatiotemporal structure of interictal discharges observed in the MEG of patients with a drug-resistant epilepsy. We consider an interictal discharge as an episode of the propagation of a cortical traveling wave.

We assume that the radial wave comes from a generating source and propagates in  $N_d^*$  different directions along the cortical surface. Bearing in mind that the distance traveled by a wave depends on the speed of its propagation, we assume that the lengths of the propagation paths of all waves are equal in terms of the number of  $N_s$  crustal nodes that the wave has visited. Thus, the  $d$ -th propagation direction can be represented

as a sequence of active cortical sources  $\mathbf{p}_d = [\mathbf{r}_d^1, \dots, \mathbf{r}_d^{N_s}]$ , where  $\mathbf{r}_i = [x_i, y_i, z_i]$  contains the coordinates of the source in the three-dimensional space,  $d \in [1, \dots, N_d^*]$ , and the first source is the same for all directions (generating source).

The activation time series of sources from the set  $\mathbf{p}_d$  form the matrix  $\mathbf{S}^d$ . To represent the spread of neural activity that generates the discharge as a wave in space and time, we model activation time series, which for each subsequent source are shifted in time relative to the previous ones. Having a direct operator  $\mathbf{G}$  with a fixed source orientation, the multichannel EEG/MEG signal,  $\mathbf{X}$ , can be represented as a linear combination of cortical traveling waves  $\mathbf{W}_d, d \text{ projected into sensor space} \in [1, \dots, N_d^*]$ :

$$\mathbf{X} = \sum_{d=1}^{N_d^*} \alpha_d \mathbf{G}_d \mathbf{S}^d + \mathbf{E} = \sum_{d=1}^{N_d^*} \alpha_d \mathbf{W}^d + \mathbf{E}$$

The matrix  $\mathbf{G}_d$  is formed from the columns of the forward operator matrix  $\mathbf{G}$ , topographies of sources from the path  $\mathbf{p}_d$ . The  $\mathbf{E}$  matrix models background brain activity and additive sensor noise. The  $\alpha_d$  coefficients correspond to the contribution of each propagation direction to the observed MEG activity.

#### 4.2.2 Basic waves

For the data model presented above, we assume that the propagation of MEG activity can be represented as a linear combination of traveling waves in sensor space. The main idea of the technique proposed here is to generate patterns of traveling waves, which we call basic waves, and then find their combination with the smallest number of terms that best explains the MEG data.

For simplicity, we define the number of active cortical sources along each propagation path as equal to the number of observations made during the event:  $N_s = T \cdot fs$ , where  $T$  is the duration of the event in seconds,  $fs$  is the frequency discretization. At the modeling stage, we consider the case when the activation time series for each of the  $N_s$  sources has a sinusoidal waveform and is shifted in time relative to the starting point. For each propagation direction, the source time series matrix  $\mathbf{S}^d$  is formed from the rows:

$$\mathbf{S}_i^d = 1 + \cos\left(\frac{2\pi(t - k_i)}{N_s}\right), k_i \in [1, \dots, N_s], t = [1, \dots, N_s]$$

The positions of the sources  $\mathbf{p}_d = [\mathbf{r}_d^1, \dots, \mathbf{r}_d^{N_s}]$  in each particular case depend on the individual anatomy, the positions of the original source  $\mathbf{v}_s = [x_s, y_s, z_s]$  and wave propagation speed. For each basic wave, we need to find a path on a graph with  $N$  vertices connected according to the  $\mathbf{A}$  adjacency matrix defined by the cortical mesh model. For a given initial position on the cortex with  $N_d$  nearest neighbors, we define  $N_d$  basis waves propagating in the directions of these nearest neighbors. For the convenience of analysis in practical applications, we do not add new vertices or edges to the graph corresponding to the cortical model. A limitation of this approach is the fact that the number of propagation directions depends on the density of vertices in the scrutinized cortical region, and also, in the case of adaptive meshes, on local curvature. The latter makes sense, since the spatial resolution of the MEG correlates with the local curvature [Nasiotis et al. \(2017\)](#).

In this paper, we describe in detail the algorithm for generating propagation paths for the starting point  $\mathbf{v}_s$ . We generate sets of basic waves for different propagation velocities: from 0.3 to 1.5 m/s [Alexander et al. \(2015\)](#).

In addition to directed radial waves, we also considered a spherical wave propagating simultaneously in all directions and consisting of a sum of radial waves, however, our tests on model and real data showed that spherical waves are usually not chosen by the algorithm as participants in the optimal combination.

When changing the propagation velocity, we also introduce a timestamp for the beginning of the wave. The exact time of wave initiation is not known, but the optimal value can be found using the sliding window method. We automatically scan the time interval containing the interictal discharge, match the basic waves to this interval, and repeat the entire analysis for the time series shifted by one time count.

### 4.2.3 Optimal combination of traveling waves

After the basic waves are formed, the next step in the analysis is to find their linear superposition that best describes the observed MEG data. Based on physiological assumptions, the desired combination should contain only a few basic waves corresponding to a few dominant propagation directions. Therefore, we are looking for the most sparse solution that describes the data and matches a small number of well-defined dominant propagation directions.

To find the contribution of each precomputed basis wave to the MEG data, we used the LASSO [Tibshirani \(1996\)](#) method, with the additional constraint that the LASSO coefficients must be positive. Since we are considering a multichannel problem, we have vectorized the data matrix  $\mathbf{X}$  and the basis waves on the sensors. The optimization problem is formalized by the equation:

$$\min_{\alpha_0, \dots, \alpha_{N_d}} \left\| \text{vec}(\mathbf{X}) - \sum_{d=0}^{N_d} \alpha_d \cdot \text{vec}(\mathbf{W}_d) \right\|^2 + \lambda \sum_{d=0}^{N_d} |\alpha_d|$$

subject to  $\alpha_d \geq 0, d = 0, \dots, N_d$

This procedure is then applied to all sets of generated basis waves with two parameters: propagation velocity and wave propagation start time. The best solution is chosen according to the  $R^2$  metric (i.e. the percentage of variance explained).

An important issue in the generation of basic waves is the discovery of the source that initiates the propagation of the wave. We determine the region of interest (ROI) as a first approximation using the RAP-MUSIC [Mosher and Leahy \(1999a\)](#) dipole fitting algorithm. To improve the accuracy of the solution, we scan the ROI using the cortex nodes that fall there as starting points and compare the solutions using the  $R^2$  metric.

### 4.2.4 Main results

We proposed a methodology for non-invasive study of the fine spatiotemporal structure of interictal discharges observed in the MEG of these patients with drug-resistant epilepsy. First, we studied the properties of the algorithm on realistic models. We then used the developed algorithm to analyze the local distribution of interictal discharges in patients with epilepsy. Preliminary results from nine patients data have shown that

the spikes coming from the epileptogenic zone show a better fit of the wave model than spikes from other regions. We believe that information on the spatiotemporal dynamics of the spread of interictal activity in the future may be useful for planning a more sparing surgical intervention.

In this work, both in model and real data, we considered only the case of focal epilepsy, assuming that the interictal discharge is generated by a well-localized area of the cortex and then propagates locally in accordance with the traveling wave mechanism. We did not consider the case of generalized interictal discharges, which often involve deeper brain structures, since it is in the scenario of focal epilepsy that the application of the developed algorithm makes practical sense, allowing additional information to be obtained for the localization of the epileptogenic zone.

#### 4.2.5 Results on model data

The Monte Carlo simulation was carried out for three signal-to-noise levels: SNR values = 1, 2, and 3. We constructed ROC curves demonstrating the ability of the proposed algorithm to detect traveling waves and distinguish them from spatially static activation. To construct these curves, 300 Monte Carlo trials were used, in which wave propagation was modeled with randomly selected propagation velocity and 300 trials, in which only static activity was simulated without propagation over cortical mantle. For the three SNR levels we have obtained ROC AUC values of 0.78, 0.95, and 0.97, which means that at a reasonably high signal-to-noise ratio, the proposed method successfully separates propagating and static activity.

Next, we evaluated the quality of reconstruction of the simulated propagation velocity. We have built a correspondence between the true model values of the propagation velocity and the values obtained by the algorithm for various SNRs. For SNR = 1, the algorithm tends to significantly overestimate the propagation velocity over compared to the true value: the reconstructed speeds do not match the true value, except for the highest propagation speed. For SNR = 2, there are still many errors in determining the speed, but the absolute difference between the estimated and actual values is much lower than for the previous case. For SNR = 3, the modal estimated value is the same as or closest to the actual speed for all but two cases where the estimated speed is overestimated. It is important to note that errors in velocity estimation are inevitable even for high SNR values due to the error that we put in when localizing the starting point, and due to the use of a more sparse shell - a model of the cerebral cortex for solving the inverse problem. Given that we consider the propagation time to be fixed, if the starting point of the wave launch found by the algorithm is shifted relative to the actual one towards the end point of the propagation path, the speed will naturally be underestimated. And vice versa, if the starting point is shifted in the opposite direction from the end of the path, then the speed will be overestimated. The higher the SNR in the data, the smaller these errors are.

We then estimated the errors in the found propagation direction. The error was calculated as  $1 - \cos(\phi)$ , where  $\phi$  is the angle between the actual and estimated main propagation directions. The values of this metric range from zero to one. For all SNR levels, most of the errors turned out to be less than 0.1, and all errors tend to decrease with increasing signal-to-noise ratio.

Although the proposed method successfully finds traveling waves and reconstructs

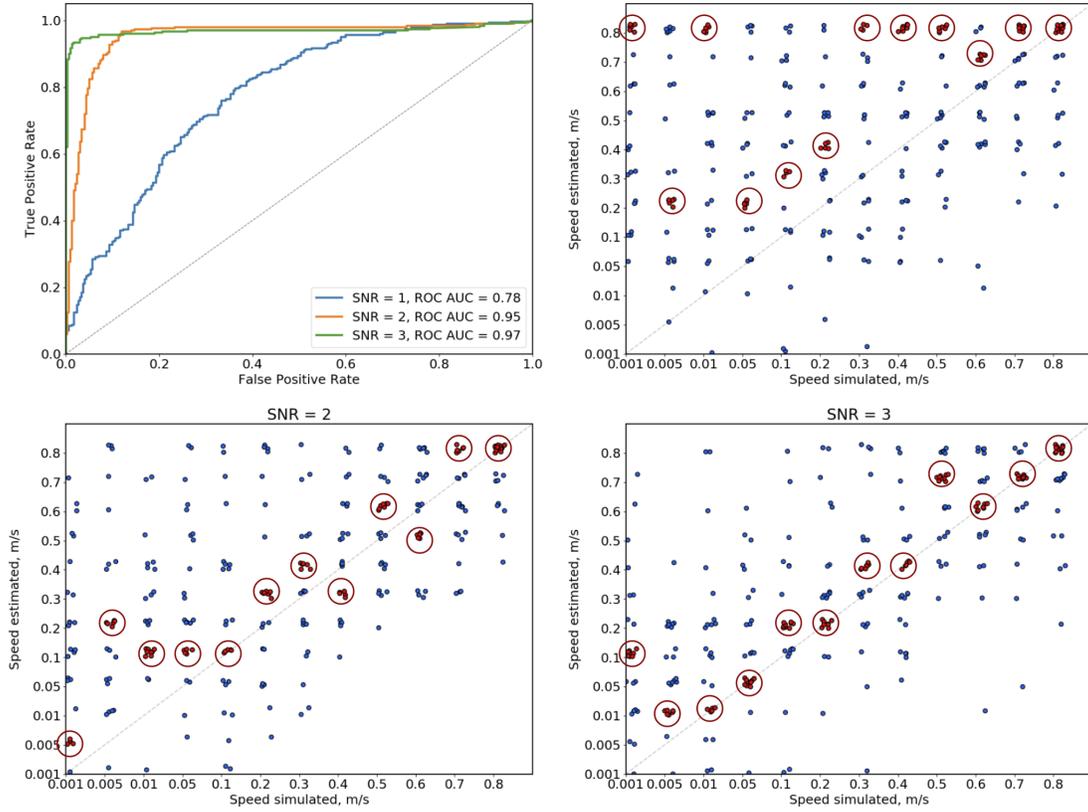


Рис. 23: Comparison of model values and their estimates for different signal-to-noise ratios.

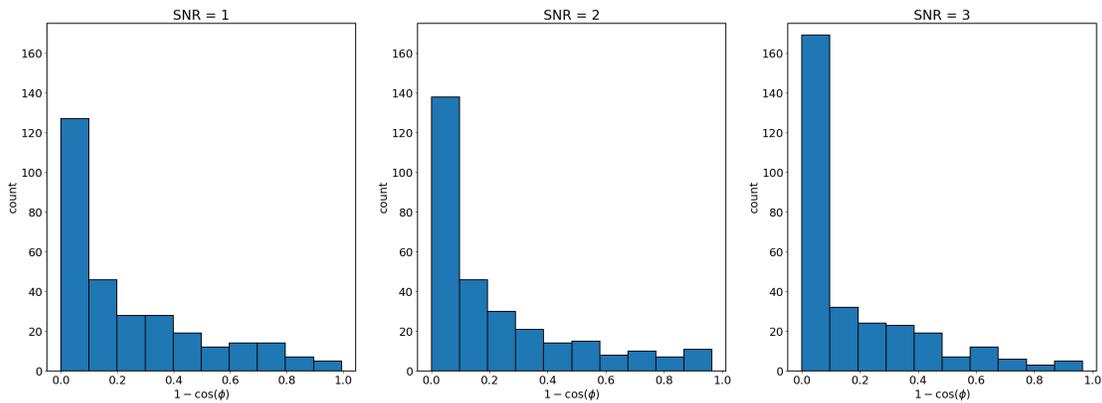


Рис. 24: Statistical distributions of propagation direction estimation error,  $1 - \cos(\phi)$ , where  $\phi$  is the angle between the main directions of actual and calculated propagation. The results are shown for 300 Monte Carlo trials and for three SNR levels: SNR is 1, 2, or 3.

their anatomical paths, it is still subject to errors due to (1) uncertainties in the wave start point estimation and (2) inaccuracies in the cortical surface parameterization. Errors resulting from the first cause can be reduced by selecting high-amplitude discharges for analysis. The second problem can be solved by performing a more accurate brain scan (7T MRI).

## 4.2.6 Results in these patients

As real data, we used data from nine patients with epilepsy: 10-minute MEG recordings during sleep. For automatic detection of interictal discharges, we used ASPIRE [Ossadtchi et al. \(2004b\)](#), which is based on the method of independent components (ICA). Then, for each of the found events, we selected the corresponding electric dipoles using the RAP-MUSIC [Mosher and Leahy \(1999a\)](#) algorithm. The localization of the sources generating the detected events on the cortex allows us to assess the physiological plausibility of the automatically detected events. We used 0.97 as a threshold for the subspace correlation metric, and all events for which RAP-MUSIC found less correlation were removed from subsequent analysis.

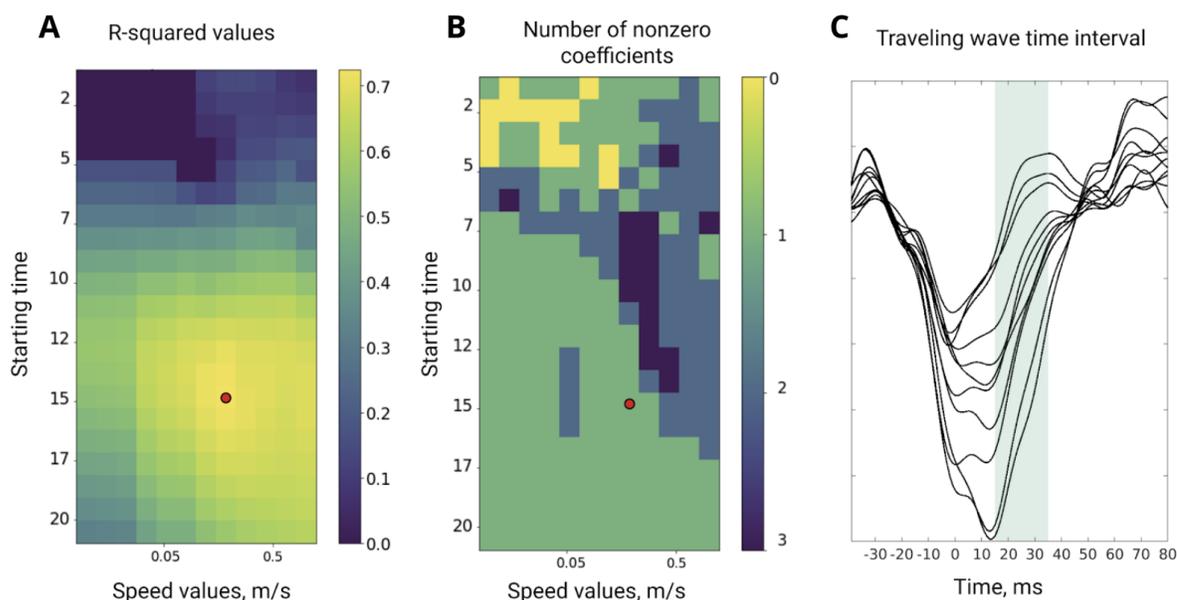


Рис. 25: A representative example of the analysis of a single interictal discharge. **A.**  $R^2$  values for various starting points and propagation speeds. **B.** Number of non-zero coefficients in the **C** model. Time series of several MEG channels containing a spike.

The proposed algorithm was applied to each found interictal discharge separately. We then applied a simple deterministic clustering algorithm based on the proximity between the obtained localizations of spikes to combine all sources into dense clusters with a radius of no more than 1 cm, each containing at least ten dipoles. ASPIRE parameters were found empirically and were fixed for all patients. Despite the fact that the described automatic detection procedure was run separately for gradiometers and magnetometers, the resulting clusters turned out to be approximately the same. All the results shown below are calculated for pre-processed magnetometer signals using MaxFilter.

We applied the proposed method to each detected interictal discharge and aggregated the resulting  $R^2$  values based on their cluster membership. Since the goal of this analysis is to find a qualitative yet simple description of interictal discharge, another important factor is the number of propagation directions in the optimal solution. Analysis of patient data revealed variability in the wave model fit depending on the specific spikes. A wave pattern with a choice of only a few dominant directions is only

suitable for some of the analyzed events. We calculated fractions of spikes with an explanatory quality of at least 0.6 for each found cluster for nine patients.

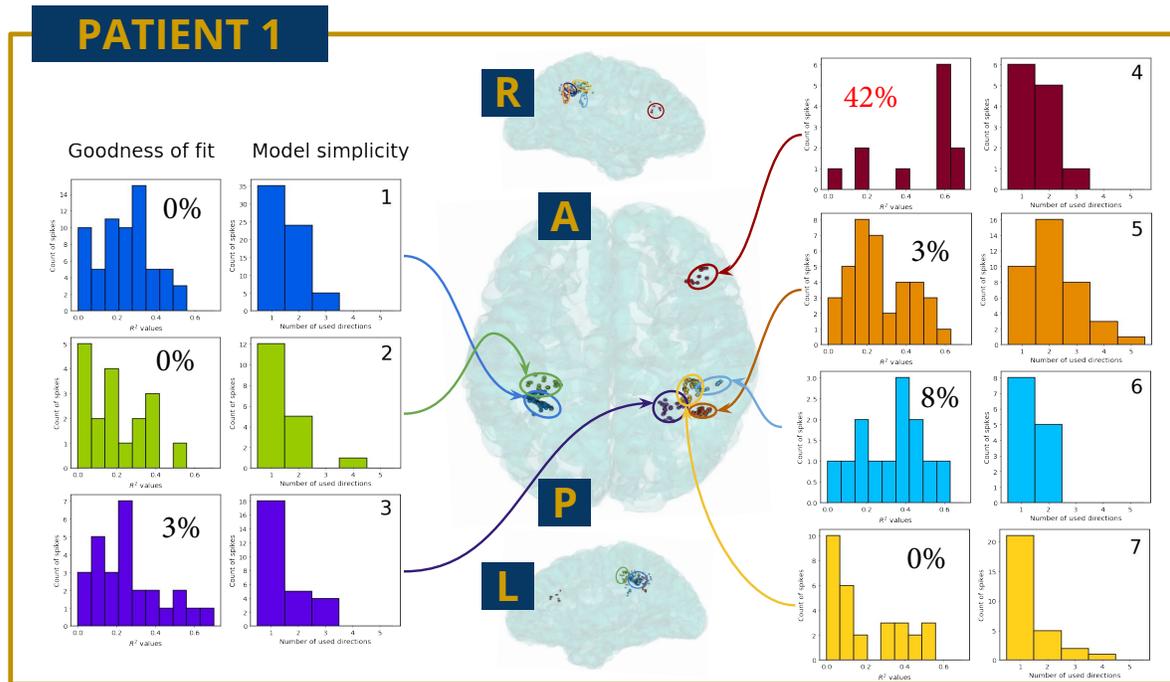


FIG. 26: Localization of seven irritative zones automatically detected by the ASPIRE technique for patient 1: axial projection from above in the middle and sagittal projection of the right and left hemispheres from above and below, respectively. Distribution of  $R^2$  metrics (degree of fit) and the number of directions used in the optimal solution (simplicity of the model) for each cluster.

Intrigued by the observed spatial segregation of traveling wave(TW)-like events, we correlated the location of the clusters with the highest percentage of spikes explained by TW and the site of resection that resulted in the Engel I outcome in each of the patients. The results are shown in the table below.

The observed wave dynamics of interictal discharges generated specifically in the epileptogenic zone, but not in other zones of irritation, is in perfect agreement with the results of other researchers who recorded brain activity using invasive methods and demonstrated that the wave dynamics is characteristic of the epileptogenic zone, for example Tomlinson et al. (2016); Diamond et al. (2021), which can be explained based on the ideas put forward in Smith et al. (2022). In our study, we independently reached the same conclusion and did so for the first time in the world based on non-invasive MEG data.

	Area of Resection / Diagnosed SOZ	Surgical Outcome	Highest $R^2$ values percentage
Patient 1 (Nmg)	rFr	I	67% (rFr)
Patient 2 (Nmg)	lParOp/sTG	n/a	70% (lParOp)
Patient 3 (Nmg)	cOcc	I	69% (cOcc)
Patient 4 (CTF)	rTmp	I	38% (rTmp)
Patient 5 (Nmg)	lTmp	I	36% (lTmp)
Patient 6 (CTF)	raTmp	I	42% (STG)
Patient 7 (CTF)	rParTmp	I	61% (rParTmp)
Patient 8 (CTF)	lmTmp	I	38% (laTmp)
Patient 9 (CTF)	laTmp	I	No clusters with $R^2 > 0.6$

Таблица 1: Information about the localization of the epileptogenic zone, the resection area and the location of the cluster with the highest  $R^2$ .

## 5 New approaches to real-time processing and decoding of brain states for neural interfaces and neurofeedback systems

### 5.1 Estimation of rhythmic activity parameters with minimal delay

An important characteristic of real-time signal processing methods is the temporal resolution and the overall delay incurred in the processing path. As described in the Introduction, the lack of temporal specificity of the feedback signal can be the reason for the low efficiency of the implementation of closed loop paradigms. Modern software solutions that implement a feedback loop and are used both in the clinic (BrainMaster, NeuroRT Training, Cygnet, etc.) and for research (OpenVibe, BCI2000) make it possible to assess the power of rhythmic brain activity with a delay exceeding 500 ms. This delay is measured from the moment data is received from the EEG device to the moment the signal is transmitted to the feedback presenting module. An additional delay of about 100 ms appears due to technical reasons, namely due to the organization of transmission between the EEG/MEG device and due to the time spent on generating the feedback stimulus by a device, such as monitor. Thus, the total delay between a cortical event (for example, a surge in the alpha rhythm) and the moment of presentation of the corresponding feedback signal in the vast majority of neurofeedback systems exceeds 600 ms. Due to the presence of such a delay, stimulation in closed-loop paradigms can occur at times that are not specific to the activity being reinforced, when the target template has already ended. An example of such activity is a surge of the alpha rhythm (8-14 Hz), which lasts about 200-300 ms. To detect and timely reinforce this kind of activity, it is necessary to reduce the delay of the closed loop system to at least 100-200 ms. In this part of the work, we describe new methods for estimating the instantaneous amplitude and phase of narrow-band signals from real-time EEG/MEG. Detailed results of the work were published in [Smetanin et al. \(2020a\)](#).

### 5.1.1 Mathematical model

On the part of the computer that receives and processes the EEG, the signal coming from the electroencephalograph is a multichannel time series with a given sampling rate, for example, equal to  $f_s = 500$  Hz. The first stage of EEG processing is the reduction of a multichannel signal to a single-channel form by means of spatial filtering. Without loss of generality, we will assume that the input to the algorithm being developed is a single-channel signal  $x[n]$ . Further, a single-channel signal  $x[n]$  can be represented as a sum of two signals:

$$x[n] = s[n] + \eta[n]$$

, where  $s[n]$  is the target narrowband signal whose power and phase are to be estimated, and  $\eta[n]$  is the broadband noise whose influence on the power estimate of interest is to be minimized. Further,  $s[n]$  can be transformed into a complex-valued analytic signal using the Hilbert transform and described as:

$$y[n] = a[n]e^{j\phi[n]}$$

where  $a[n]$  is the instantaneous amplitude of the signal (the square root of the instantaneous power),  $\phi[n]$  is the instantaneous phase,  $j$  is the imaginary unit. The estimate of  $a[n]$  and  $\phi[n]$  with known values of  $y[n]$  is obtained by calculating the absolute value and angle of the complex number  $y[n]$ , namely:

$$a[n] = \left( \Re(y[n])^2 + \Im(y[n])^2 \right)^{\frac{1}{2}}$$

$$\phi[n] = \arctg \left( \Im(y[n]) / \Re(y[n]) \right)$$

,  $\Re(y[n])$  - real part of  $y[n]$ ,  $\Im(y[n])$  - imaginary part of  $y[n]$ .

It should be noted that the operations of calculating the modulus and argument of the analytical signal do not introduce additional fundamental delay into the processing path, since they are calculated for each moment of time using the signal values only at this moment in time. However, calculating the Hilbert transform at point  $n'$  ideally requires an infinite window centered around point  $n$  on the time axis. Using the classical approximation of filters with an infinite impulse response, it is possible to represent the calculation of the Hilbert transform as the convolution of the signal and the finite impulse response of the Hilbert filter. However, such an approximation involves non-causal processing, requiring knowledge of the input values  $s[n]$  for  $n$  to evaluate the transformed signal at time  $n'$  both from the past with respect to the current time value ( $n < n'$ ) and values from the future ( $n > n'$ ). Such conversion cannot be performed in real time. However, the application of such an algorithm for a known complete EEG record allows us to extract the envelope  $a[n]$  and phase  $\phi[n]$ , which will be called the ideal envelope and ideal phase, except for the values at the edges of the corresponding record.

Thus, the problem of this section is formulated as the construction of a causal algorithm that estimates the amplitude  $a[n]$  and the phase  $\phi[n]$  in real time using a single-channel signal  $x[n]$ . In this case, the explicitly specified method delay  $D$  should be one of the parameters of the method being developed. The best method is that which

allows, with decreasing parameter  $D$ , to estimate  $a[n]$  and  $\phi[n]$  in real time with the best quality, which will be defined below.

### 5.1.2 Existing methods

The classical method (hereinafter referred to as *rect*) of extracting instantaneous power is the method based on amplitude signal demodulation and is similar to the principle of operation of a simple radio receiver that detects an amplitude modulated signal. This method includes three successive steps: narrow-band filtering in a given range, narrow-band signal 'rectification' (absolute value calculation), and signal smoothing with a low-pass filter (LPF). The output of this algorithm is an estimate of the instantaneous amplitude  $a[n]$ . The delay of this algorithm is the sum of the delays of the narrow band filter and the low pass filter. In the case when symmetrical filters with finite impulse response (FIR) are used as filters, this delay is half the sum of the lengths of the impulse responses of the filters. In this paper, FIR filters are used as filters, and the length  $N_1$  of the impulse response of the narrow-band filter and the delay value  $D$  are used as parameters. Accordingly, the length  $N_2$  of the LPF impulse response can be determined from the values of the desired total delay  $D$  and the FIR length of the first filter  $N_1$  as  $N_2 = 2D - N_1$ .

The second method (hereinafter referred to as *hilb*) is based on the windowed Hilbert transform. Signal processing in this case is performed locally using the sliding window of length  $N_3$ . The last sample inside the window corresponds to the last received sample  $x[n]$ . For each new window, the recorded signal is filtered in a narrow-band range with zero phase, then, on this window, the narrow-band signal is reduced to an analytical form using the Hilbert transform. The absolute value of the analytical signal at a point separated from the end of the window by  $D$  samples is an estimate of the instantaneous amplitude with a delay of  $D$  counts. Thus, the parameters of the method are the quantities  $D$  and  $N_3$ . It should be noted that this method is affected by the transients at the window boundary from the side of the last received data sample.

There are also several modifications of the method described above, which are actively used in closed-loop paradigms when there is a need for accurate real-time phase estimation [Schaworonkow et al. \(2018\)](#); [Bergmann et al. \(2016\)](#). One of such methods is the method based on autoregressive correction of boundary effects, proposed in [Chen et al. \(2011\)](#). This method, denoted below as *ffiltar*, is designed to estimate the phase with zero delay  $D = 0$ . In our work, for comparison, we also used the estimate of the envelope at the time  $D = 0$ . Also, methods similar to the *ffiltar* method are currently appearing, but using more complex predictive models. Examples of such methods are [McIntosh and Sajda \(2020\)](#); [Shakeel et al. \(2020\)](#).

### 5.1.3 Description of the developed family of methods

The new algorithm for estimating the parameters of the rhythmic activity of the brain proposed in this paper is based on the following idea. Let  $f_s$  denote the sample rate of the recorder. The transition from a wideband signal  $x[n]$  to an analytical narrowband signal  $y[n]$  with an additional delay of  $D$  samples can be represented as a linear stationary system with a complex frequency response (CFC) specified on the interval from  $-\pi$  to  $\pi$ , such that for frequencies  $\omega$  from  $2\pi f_1/f_s$  to  $2\pi f_2/f_s$  it takes values equal to  $e^{-j\omega D}$  and

equals 0 outside this range, including for the negative frequency range from  $-2\pi f_2/f_s$  to  $-2\pi f_1/f_s$ . The frequencies  $f_1$  and  $f_2$  are specified in Hertz (Hz) and determine the width of the spectrum of the narrowband signal. This system will be referred to below as the ideal detector of a narrow-band analytical signal with a delay  $D$ .

This system can be approximated by a causal FIR system. For this, it seems possible to use the criterion of the minimum sum of squares of the difference between the CFC of the ideal detector of a narrow-band analytical signal with a delay  $D$  and its causal approximation of finite length. The solution of the optimization problem leads to the following statement: the FIR of the approximating system  $b[n]$  is obtained from the ideal CFC using the inverse discrete Fourier transform. The parameters of the method are  $N_t$  - the length of the FIR and  $N_f$  - the number of discrete frequencies in the Fourier transform. If  $N_f > N_t$  then  $x[n]$  is padded with zeros. For non-negative delays  $D$ , such a solution, given the proper formulation of the *hilb* method, coincides with the proposed method. Negative delays make it possible to predict the signal into the future by  $|D|$  samples. As a result, the estimate of the analytical signal is obtained by convolution  $y[n] = b[n] * x[n]$ . The absolute value and angle of the resulting complex-valued signal is an estimate of the instantaneous amplitude and phase of the desired narrowband signal. The resulting envelope and phase detector is denoted as *cfir*.

Several modifications of the optimization problem can be proposed, which increase the accuracy of the designed narrow-band signal envelope detector. For example, it is possible to take into account individual spectral features. Adding the amplitude spectrum of the signal to the optimization problem as weights makes it possible to formulate the functional in accordance with the criterion of the weighted sum of squares of the mismatch and obtain an individual filter for each subject. This method is designated as the *wcfir* method. It is also possible to search for filter coefficients using an optimization problem formulated in the time domain, which makes it possible to take into account the non-stationarity of the signal of neuronal activity and use adaptive approaches based on the recursive least squares (RLS) method. The time domain adaptive method is denoted as *tcfir*. It should be noted that the *wcfir* and *tcfir* methods require prior recording of a small EEG/MEG segment to determine the signal spectrum and adjust the filter parameters. A detailed description of the developed algorithms is given in [Smetanin et al. \(2020a\)](#) and the appendix.

#### 5.1.4 Method comparison

The following metrics were used to check the quality of the developed algorithms and compare them with existing approaches. To estimate the quality of the reconstruction of the  $a[n]$  envelope for the delay  $D$ , using the correlation coefficient, we compared the estimate  $\hat{a}[n]$  and the ideal envelope  $a[n]$  shifted by  $D$  counts.

$$r_a = \frac{\sum_{n \in \mathcal{N}_a} (a[n - D] - m_a)(\hat{a}[n] - m_{\hat{a}})}{\sqrt{\sum_{n \in \mathcal{N}_a} (a[n - D] - m_a)^2} \sqrt{\sum_{n \in \mathcal{N}_a} (\hat{a}[n] - m_{\hat{a}})^2}} \quad (21)$$

Similarly, to estimate the quality of the phase reconstruction, we calculated the shift  $b_\phi$  and the root-mean-square deviation  $\sigma_\phi$  of the phase estimate  $\hat{\phi}[n]$  relative to

the ideal phase at time points  $\mathcal{N}_\phi = \{n : n \in \mathcal{N}_a, \text{sign}(\hat{\phi}[n]) > \text{sign}(\hat{\phi}[n-1])\}$  when  $\hat{\phi}[n]$  crossed the value 0 (zero phase detection):

$$b_\phi = \frac{1}{|\mathcal{N}_\phi|} \sum_{n \in \mathcal{N}_\phi} \phi[nD] \quad (22)$$

$$\sigma_\phi = \sqrt{\frac{1}{|\mathcal{N}_\phi| - 1} \sum_{n \in \mathcal{N}_\phi} (\phi[n - D] - b_\phi)^2} \quad (23)$$

We tested the algorithms using 2 minute long EEG data recorded from 10 subjects at rest with their eyes open. The recording was carried out on 32 channels of the standard scheme 10-20 with references A1-A2 and a sampling rate of 500 Hz using a NVX - 136 electroencephalograph (Medical Computer Systems LLC). Only channel P4 was used for analysis. The true envelope was estimated as the envelope of the rhythm non-causally filtered in the range of 8-12 Hz.

The performance of the algorithm was evaluated separately for each record, as well as separately for each delay  $D$  from the set of values of  $D$  from -100 to 250 ms with a step of 50 ms. Each entry was divided into two parts of 2 minutes each. For the first half of the record for each algorithm, a search was made for the parameters at which the maximum value of each of the studied metrics is reached. For the found optimal parameters, the correlation value was calculated in the second half of the record. The latter value was used as an assessment of the quality of the algorithms. This approach makes it possible to guarantee for each of the methods that the optimal parameters are chosen among the possible values, and allows for evaluation the quality on an independent sample.

Figure 27 shows the results of the evaluation of the algorithms' quality. For each value, a 95% confidence interval is specified, calculated using the bootstrap method with 1000 iterations according to the statistics "average value over a sample of EEG records".

As expected, the accuracy of the envelope estimation (Fig. 27A) improves as the delay parameter  $D$  increases. The *rect* method exhibits the fastest decay in envelope estimation quality with decreasing delay parameter. For delays less than 150 ms, it becomes difficult to use this method due to the unsatisfactory quality of envelope reconstruction. The family of methods developed in this work makes it possible to better preserve quality with a decrease in delay, while the *wcfir* method demonstrates the best result at each point. The envelope estimation for the *ffiltar* method by construction is available only for a delay equal to zero. The quality of the envelope at this point is comparable to complex-valued filters. However, it should be noted that this method requires calculating the parameters of the AR model at each step and setting additional parameters, the optimal values of which may change during the experiment. Thus, this method is more time-consuming to use, more expensive in terms of the number of calculations, depends on a large number of parameters and does not allow us to adjust the delay parameter, which significantly complicates the use of this approach in closed-loop paradigms that require a quick assessment of the parameters of the rhythmic activity of the brain..

The phase recovery accuracy metrics are presented in panels B, C, and D. For non-negative delays, the  $b_\phi$  offset and the absolute value of the offset behave similarly for

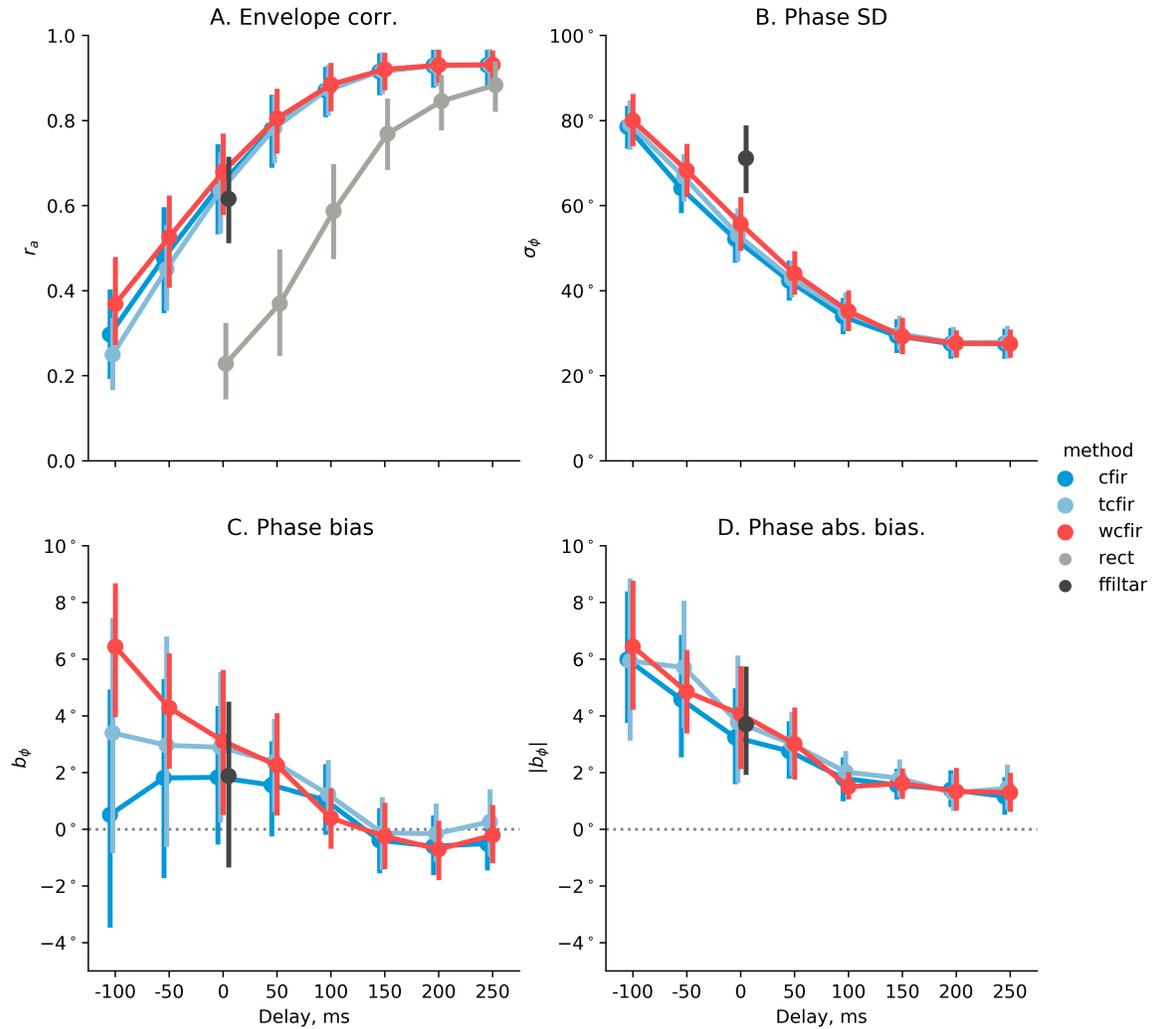


Рис. 27: Envelope and phase recovery quality metrics for different methods and at different delays

all methods and do not exceed 5 degrees for all methods. As for the phase standard deviation, complex-valued filters show the best value of this metric as compared to the other approaches.

An analysis was also made of the influence of the signal-to-noise ratio(SNR) on the accuracy of the envelope and phase reconstruction. A detailed description of the results can be found in [Smetanin et al. \(2020a\)](#). Here we only note that the quality of reconstruction improves with increasing SNR for all methods. At the same time, the newly developed methods demonstrate the greatest resistance to noise.

In addition, an analysis was made of the applicability of the developed approaches in discrete paradigms, where the start of stimulation of brain activity (for example, using transcranial magnetic stimulation) is tied to the moments when a certain envelope threshold is exceeded, for example, the 95% percentile. The analysis showed that for zero delay, the developed methods allow achieving 75% accuracy in detecting such excesses.

## 5.2 Decoding brain activity using interpretable neural networks

### 5.2.1 Introduction

Brain-computer interfaces decode information from neural activity and generate control signals for external devices. The use of deep learning techniques to decipher brain activity makes it possible to replace the traditional manual approach to identifying informative features with an automatic procedure performed by the first few layers of a deep neural network, the parameters of which are tuned in the context of the specific classification task. Taking into account the physiology of the recorded brain activity signals and the physical properties of the method their registered with makes it possible to formulate and impose restrictions on the architecture of the first layers of deep neural networks, which ensure the interpretability of network’s weight coefficients. In the case when results of such an interpretation reflect the physiologically expected components of the recorded signals (spatial distribution, frequency range), we can count on the reliability of the obtained decision rules and expect a certain degree of generalization. In addition, physiologically meaningful results of interpretation often make it possible to draw a conclusion about the sources of information used by the network to solve the assigned classification problem. This often makes it possible to identify undesirable situations when, for example, muscular activity is used as a useful signal instead of brain activity. In addition, such interpretability of the learned decision rules opens up exciting opportunities for building automatic systems for extracting new knowledge (knowledge mining) from the data of standard experiments and searching for patterns that go beyond the traditional approach, that consists in quantifying additive changes in the activity of certain areas of the cerebral cortex that accompany the presence of the studied cognitive process. For these purposes, we have developed a new compact architecture of a convolutional neural network and a theoretically justified approach to interpreting the weight coefficients of the first two layers of the network. This allows us to estimate the localization and dynamic characteristics of neuronal populations, whose activity appear pivotal in solving the classification or decoding problem at hand. Our new weights analysis method naturally extends the classical approach to interpreting the parameters of regression models known from estimation theory and adapts it to the situation when spatial and frequency filtering coefficients are simultaneously adjusted as part of the learning process. The spatial and frequency profiles obtained as a result of this interpretation characterize populations of neurons that are of key importance for a particular decoding task and can be used for subsequent analysis by fitting electromagnetic and dynamic models that describe the geometric and frequency domain properties of the corresponding neuronal populations.

### 5.2.2 Signal Observation Model

We traditionally start with an observed signal model, which is a slight modification of the observation model used in the previous sections. The purpose of this modification is to highlight the presence in the data of signals from neuronal populations unrelated to the task, the so-called “brain noise”. As shown in 28, the signal vector  $\mathbf{x}[n]$ , measured using an array of sensors, is the result of the activity of many neural populations, the activity of one part of which ( $G_i$ ) is associated with some behavioral variable  $z[t]$ , such as, for example, hand kinematics, and the activity of neural populations, denoted by

$A_i$ , is not related to this behavior to be decoded. The activity of neuronal populations is

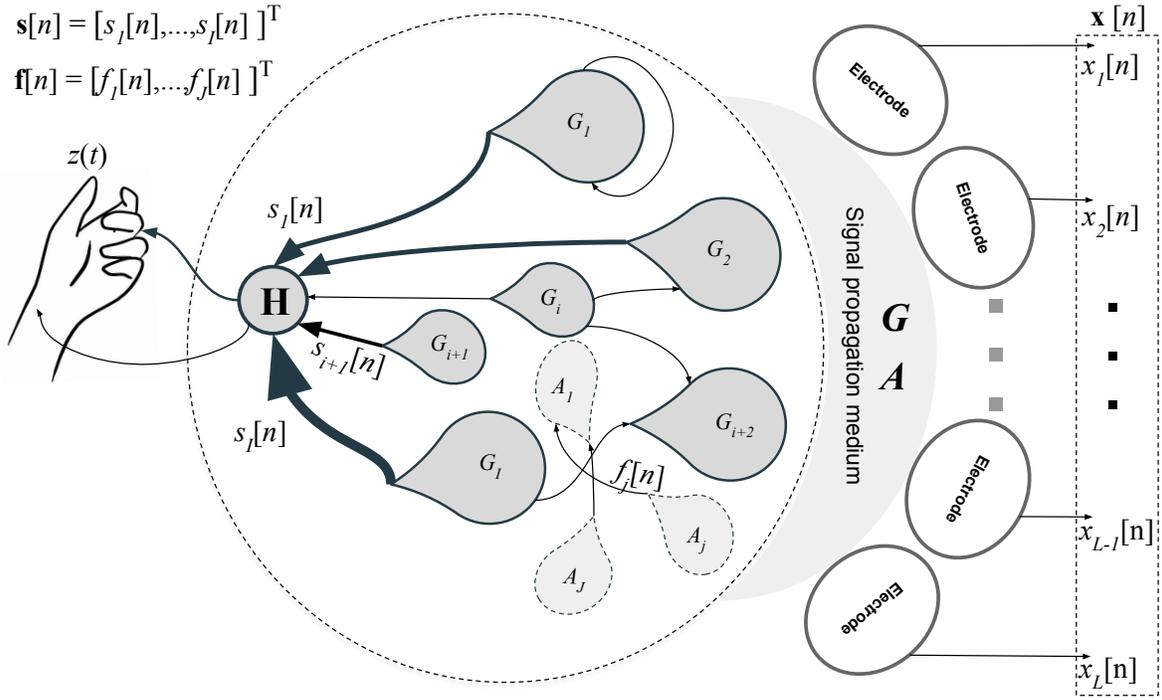


Рис. 28: Phenomenological model

related in a complex non-linear way to the decoded behavior, however, there is a strong belief that the decoded behavior depends on the instantaneous intensity of the rhythmic activity of spatially segregated populations of neurons, and thus the key information is contained in the readings of the rhythmic activity envelope extracted from the observed mixture of signals using the operation of spatial filtering [Sabbagh et al. \(2020\)](#). It is important to emphasize that, in accordance with modern concepts, the signals recorded using EEG or ECoG are linear superposition of fluctuations of the local field potentials  $s_i[n]$  and  $f_i[n]$ , rather than their envelopes. Thus, taking into account the properties of instantaneous propagation of electromagnetic activity described earlier, the observation equation for  $\mathbf{x}[n]$  can be written as:

$$\mathbf{x}[n] = \mathbf{G}\mathbf{s}[n] + \mathbf{A}\mathbf{f}[n] = \sum_{i=1}^I \mathbf{g}_i s_i[n] + \sum_{j=1}^J \mathbf{a}_j f_j[n] = \sum_{i=1}^I \mathbf{g}_i s_i[n] + \eta[n] \quad (24)$$

Here  $\mathbf{g}_i$  and  $\mathbf{a}_i$  are  $N_{sens} \times 1$  vectors of target and non-target source topographies. It is important not to forget that  $s_i[n]$  activities are characterized by a certain frequency range containing information about the process being decoded, and the  $f_i[n]$  signals actually make the observed mixture of target signals noisy. The amount of noise in different frequency ranges depends on the ratio of the frequency characteristics  $s_i[n]$  and  $f_i[n]$ .

In the simplest case, when  $I = 1$  and  $J = 1$ , that is, when there is one target and one non-target signal in the mixture, the task of the algorithm for decoding the behavior of  $z(t)$  from neuronal activity is reduced, on the one hand, to setting such spatial filter  $\mathbf{w}$ , such that the signal  $\mathbf{w}^T \mathbf{g}_1 = 1$  and  $\mathbf{w}^T \mathbf{a}_1 \rightarrow \min$ , and on the other hand, the

search for such a frequency filter  $H(f)$ , which, in accordance with the matched Wiener filtering recipe, would “collect” information from all frequency ranges in which there is information about  $z(t)$  with weights inversely proportional to the degree of noise in these frequency ranges. At the same time, it is necessary to understand that the *spatial* filter  $\mathbf{w}$  is tuned to the target source and simultaneously detuned from noise in the context of the *frequency* filter  $H(f)$  and vice versa. Thus, the spatial profile  $\mathbf{g}_1$  of the target source and the power spectral density (PSD or frequency profile) of its activity ( $S(f) = PSD(s[n])$ ) turn out to be non-trivially related to the vector of coefficients spatial filter  $\mathbf{w}$ , frequency filter characteristic  $H(f)$ , topography of the source of non-target activity  $\mathbf{a}_1$  and its PSD ( $F(f) = PSD(f[n])$ ).

### 5.2.3 Compact Convolutional Network Architecture

Our architecture for the general case is shown in Figure 29 and is a set of stereotypical processing branches, each of which contains a combination of spatial and frequency filters that hypothetically extract a signal from a certain neuron population, followed by the calculation of the envelope of this activity using  $\text{RELU}(-1)$  and smoothing with one more convolution along the “time” axis. Let us pay attention to the fact that the operation of spatial filtering, which weights signals from different sensors with certain coefficients, and frequency filtering, implemented by convolution along the “time” dimension of the input data matrix, are factorized in our architecture, which ensures the necessary physiology of the subsequent interpretation. Further, the output signals of several such branches, hypothetically representing the envelopes of the activity of individual neural populations in the most informative frequency ranges, are fed to the input of a fully connected neural network, whose architecture may include more complex layers and components, such as LSTM, as we do in the decoding task speech, described below, as well as various options of transformers, etc.

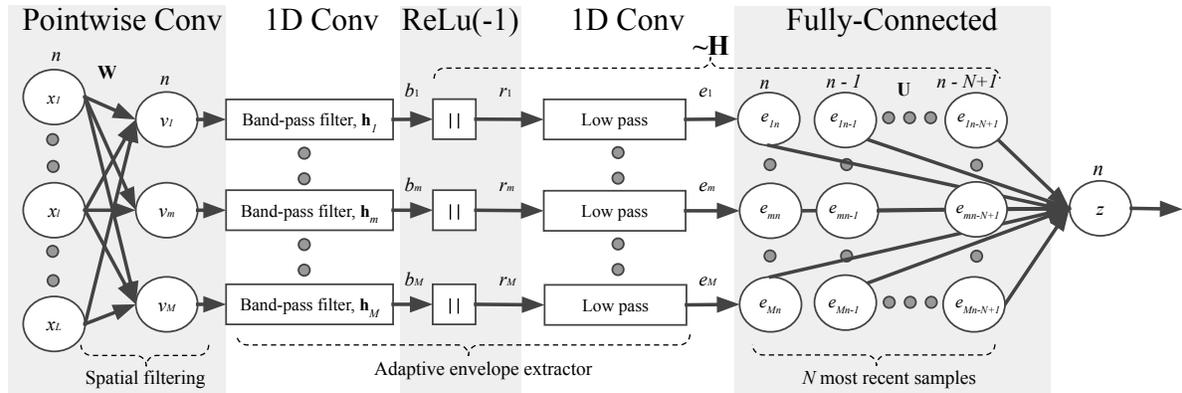


Рис. 29: The compact convolutional network architecture includes several branches - an adaptive envelope detector, receiving spatially unmixed inputs, and outputting envelopes whose latest  $N$  values are indexed  $n - N + 1, \dots, n$  are converted to the decoded variable  $z(n)$  by a fully connected layer.

### 5.2.4 Weights interpretation method

Consider the first two layers of the architecture shown in Figure 29 and the input data matrix  $\mathbf{X}[n] = [\mathbf{x}[n - T + 1], \dots, \mathit{mathbf{x}}[n]]$ , provided that the length of the filter of the second layer, that performs the frequency filtering operation, is equal to  $T$  samples. Then the result of processing the input data using the spatial  $\mathbf{w}_i^*$  and the frequency  $\mathbf{h}_i^*$  filters of the  $i$ -th branch can be written compactly as

$$b_i^*[n] = \mathbf{w}_i^{*T} \mathbf{X}[n] \mathbf{h}_i^* \quad (25)$$

Here  $*$  denotes the optimal values of the weight vectors obtained as a result of training. At the output, we have some hypothetical target narrow-band process  $b_i^*[n]$ , reflecting the activity of the neural population, the envelope of which is subsequently used by the neural network in the final decision rule. We assume that both filters, as a result of training, are tuned to extract the sequence  $b_i^*[n]$ , which best approximates some unknown signal of neuronal population activity in the least squares sense.

Consider the perturbation around the optimum ( $\mathbf{w}_i^*, \mathit{mathbf{h}}_i^*$ ) of the spatial filter coefficients  $\mathbf{w}_i = \mathbf{w}_i^* + \epsilon$ . As a result, we will have a simple regression modeling problem:

$$\mathbf{w}_i^* = \mathit{argmin} E\{ \|\mathbf{w}_i^T \mathbf{X}[n] \mathbf{h}_i^* - b_i^*[n]\|^2 \} = \mathit{argmin} E\{ \|\mathbf{w}_i^T \mathbf{z}[n] - b_i^*[n]\|^2 \} \quad (26)$$

Assuming the Wiener optimality of the found solution, the independence of  $b_i^*[n]$  for different branches, and using the principle of orthogonality, it is easy to show that

$$\hat{\mathbf{g}}_i = \sigma_{b_i}^{-2} E\{ \mathbf{z}[n] \mathbf{z}^T[n] \} \mathbf{w}_i^* \quad (27)$$

where  $\hat{\mathbf{g}}_i$  is the estimate of the spatial profile (topography) of the  $i$ -th source, and  $\sigma_{b_i}$  is the standard deviation of  $b_i[n]$ . Thus, through learning the decoding task, we gain access to the spatial activity profiles of the key neural populations, to which the branches of our architecture are tuned. It is important to answer that when calculating the estimates of the topography of the neural population to which the  $i$ -th branch is tuned, we used the spatial correlation matrix of data filtered by the frequency filter  $\mathbf{X}[n] \mathbf{h}_i^*$  of this very branch. Thus, the frequency context is taken into account, in which the spatial filter  $\mathbf{w}_i$  is configured. The influence of the spatial structure characteristic of non-target activity is taken into account through the covariance matrix of data  $\mathbf{z}[n]$ , passed through the corresponding spatial filter.

In addition to the geometric characteristics of neuronal populations, the frequency properties of their activity are of interest. As can be seen from 25, spatial and frequency-selective filtering are algebraically identical, and therefore, using similar perturbation analysis, while weakening the requirement for the equality of the length of the data segment and the vector of filter coefficients, replacing the product with a convolution and going into the frequency domain, it is easy to show that the power spectral density (or frequency profile)  $Q_i(f)$  of the activity of the target population can be calculated as

$$Q_i^*(f) = P_i^*(f) H_i^*(f) \quad (28)$$

Here  $P_i^*(f)$  is the power spectral density of the spatially filtered input data using the spatial filter of the  $i$ -th branch, and  $H_i^*(f)$  is the Fourier transform of the weights of

the second temporal convolution of the  $i$ -th branches of our architecture. A similar conclusion could be reached using a more classical approach based on the expression for the frequency response of a matched Wiener filter operating under conditions of additive noise independent of the target signal, which is described in [Petrosyan et al. \(2020a\)](#). However, the approach described here makes it easy to see a parallel with the expression for interpreting the spatial weights [27](#), since the power spectral density is actually the Fourier image of the autocovariance sequence of a random process.

It is important to contrast the method proposed here with a more common approach, in which, for example, when interpreting the weight coefficients of spatial filtering, the [Haufe et al. \(2014\)](#) rule for regression models is directly transferred without taking into account the presence of a frequency filter in the corresponding branch. Also, we would like to emphasize that for the first time we introduced the concept of the frequency profile of the population activity, which describes the power spectral density of the entire dynamics of the neuronal population, and not just the frequency range that turned out to be the least noisy and tuned away from the activity of other groups of neurons that are not related to the decoded process.

It is especially important to study the combination of spatial and frequency profiles, since it allows one to judge about the neuronal origin of the signals used. The fact is that neuron populations have a characteristic dependence that links the population size with the central frequency of its activity [Volkova et al. \(2019a\)](#); [Müller et al. \(2004\)](#). The more spatially distributed the population is, the lower the characteristic frequency of its activity and vice versa. On the other hand, in the case of ECoG analysis, muscle activity is characterized by spatially spread out activity due to the fact that signals from scalp muscles undergo diffusion due to propagation through low-conductivity tissues of the skull. In this case, the frequency range of myographic activity corresponds to 30-250 Hz and lies to the right of the traditional frequency range of the activity of neuronal populations. Accordingly, if a branch of our architecture is tuned to a muscle source, then a joint analysis of the spatial and frequency profiles should unequivocally identify this based on a combination of high characteristic frequency and spatially non-focal distribution. Conversely, the presence of an inverse relationship, in which reduction in the spatial size of the population leads to an increase in the characteristic frequency of activation, should serve as an indicator of the neuronal origin of information sources to which the neural network is tuned.

### 5.2.5 Results

To test the efficiency of the proposed approach, we carried out a number of numerical experiments using the numerical simulation technique. Details and results of such modeling are described in [Petrosyan et al. \(2020a\)](#). As part of these experiments, we studied various scenarios, combinations of target and non-target neuronal sources, and noise levels. In all cases, we demonstrated a good agreement between the estimates of source profiles and those that were included in the simulation. [Figure 30](#) shows the results of a Monte Carlo study in which the spatial and frequency profiles were generated randomly. Target variable  $z(t)$  was generated as a linear combination of envelopes of a subset of sources, linear combination coefficients were generated randomly for each Monte Carlo iteration. Since we are primarily interested in the shape of the profiles, and not their absolute values, we used the correlation coefficient between the restored

and the original profiles as a metric.

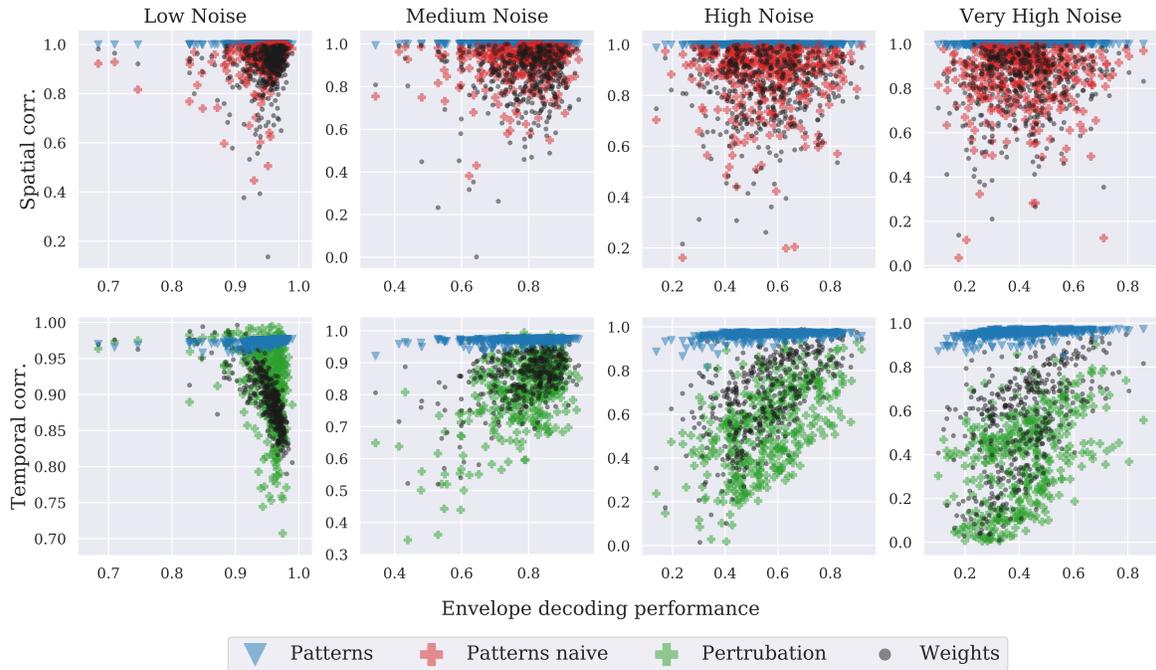


Рис. 30: Monte Carlo simulation. The point coordinates reflect the envelope decoding performance achieved (x-axis) and the true profile correlation coefficient (y-axis) for each Monte Carlo test. Each color-coded point corresponds to one Monte Carlo test and encodes the method used to calculate the profiles. *Weights* - direct interpretation of weights, *Patterns naive* - interpretation of spatial profiles without considering branch-specific temporal filters, *Patterns* - suggested method.

As can be seen from Figure 30, in all cases, the proposed technique (blue markers) provides high accuracy in the reconstruction of both frequency and spatial profiles of hidden sources. In the comparative analysis, we used the method of analyzing the significance of features based on the analysis of perturbations (green marker) and the traditional technique adopted in the “neuroimaging” community without taking into account the contribution of the corresponding frequency or spatial filter (red marker). It is noteworthy that for a situation with a high noise level (the rightmost column), even with a low decoding quality, the interpretation accuracy was preserved, which indirectly indicates the unbiasedness of the estimates obtained.

### 5.2.6 Decoding finger kinematics from ECoG

By applying this architecture first to the test ECoG data publicly available via Berlin BCI Competition IV website and to the task of decoding finger movement, we achieved a quality that is statistically indistinguishable from the best result of the competition, but without the use of manual generation of informative features. The details of this experiment are described in [Petrosyan et al. \(2020a\)](#). This public data did not contain information about the position of the electrodes, which made it impossible to visualize the results of the spatial structure interpretation. To test our new method, we applied it to the ECoG data that we registered as part of a project to develop the algorithmic

foundations of bidirectional invasive neural interfaces at HSE’s CBI. The results are shown in figure 31.

The spatial profile is depicted as an 8 x 8 rectangle corresponding to the shape of the electrode grid located on the patient’s sensorimotor cortex. The uppermost spatial profile is concentrated on only a few electrodes and corresponds to a frequency profile (blue curve) with a maximum around 75 Hz, in the high gamma range and could be a manifestation of muscle activity, if not for the corresponding compact (illuminating only two electrodes) spatial profile. The middle group of profiles is characterized by a lower characteristic frequency and, accordingly, a significantly more spatially distributed representation on the grid, which is what we expect in the case of the neuronal origin of the corresponding information source Müller et al. (2004); Volkova et al. (2019b). And finally, an even lower frequency profile with even more spatially distributed spatial activation profile.

Next, using the well-known phenomenon of the somatotopic organization of the sensorimotor cortex as an example, we will illustrate the idea of automatic knowledge mining. After solving the problem of decoding the movement of each of the five fingers, for each of them we chose the most significant branch (the removal of which led to the largest decrease in decoding accuracy) and displayed spatial profiles taking into account the position of the electrode grid on the patient’s cortex. This information was obtained by comparing post-implantation CT with the patient’s cortex extracted from volume MRI using the FreeSurfer software.

The topographies of sources given in the left panel correlate well with the existing knowledge about the somatotopy of the sensorimotor homunculus depicted in the right in Figure 31. The thumb has a representation on the cortex, which occupies the maximum area in comparison with the zones of representation of the other fingers and is located below the zones of their representation. The zone of representation of the index finger is located below the zone of the middle finger, which in turn is located below the zone of the little finger. Thus, having solved the problem of decoding the kinematics of fingers using an interpretable deep neural network and interpreting the weight coefficients of spatial filters, we automatically obtained correct knowledge about the somatotopic organization of the sensorimotor cortex.

### 5.2.7 Speech decoding from invasive data

Methods for decoding speech from brain activity will soon form the basis of neural interfaces capable of restoring a key human function in case of its loss due to trauma, surgery, or neurodegenerative disorders. The functioning of this device is based on a mathematical algorithm capable of decoding the activity of specific areas of the cerebral cortex in which the speech function is represented, thus generating control commands for the speech synthesis device. While such a solution seems ideal, practically implemented approaches rely on decoding a discrete set of words from the activity of electrocorticographic electrodes distributed over a large area of the cerebral cortex. Implanting such grids requires full-scale neurosurgical intervention and carries significant risks for the patient. Therefore, practically applicable systems should be based on relatively compact electrode grids, whose implantation is less risky. Thus, the development of speech neural interfaces based on the registration of brain activity using a small number of compactly located electrodes seems to be relevant.

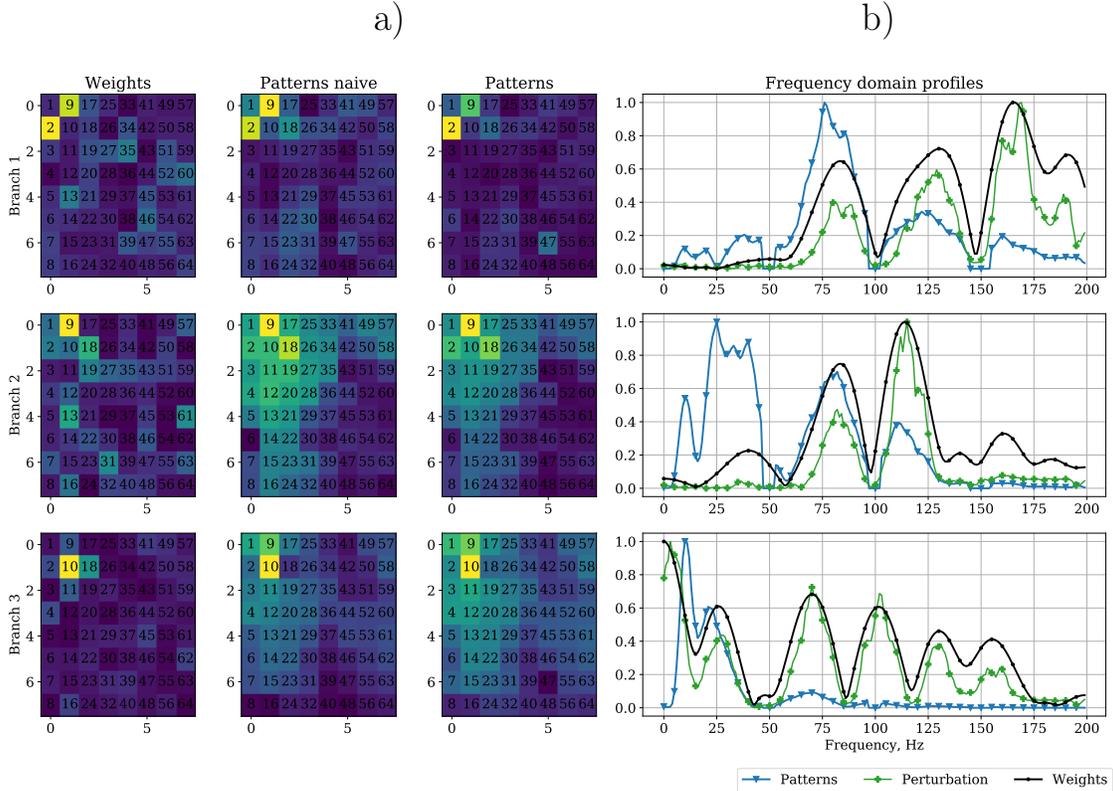


FIG. 31: Interpretation of network weights for the decoder of little finger kinematics in patient CBI 2 (ECOG). Each line of the graphs corresponds to one of the three branches of the trained decoder. a) The leftmost column shows the color-coded spatial filter weights, the next two columns correspond to the naively and correctly reconstructed spatial patterns. Blue corresponds to the minimum absolute activation, and yellow corresponds to the maximum. b) Interpretation of the temporal convolution weights in the Fourier domain. FFT filter weights - (black  $\bullet$ ), power spectral density (PSD)  $Q_m^*[k]$  of LFP pattern (blue  $\blacktriangledown$ ) obtained according to equation (28). The other line (red diamond) is the power spectral density of the signal at the output of the temporal convolution block. The results of sensitivity analysis using the perturbation approach are shown in (green +)

Petrosyan et al. (2021b) describes our experiment on recording brain activity in patients during a speech task and developing a method for decoding speech based on brain activity signals recorded using a small number of compactly located invasive sensors. As part of this study, we completed the entire cycle of work from the development of an experiment paradigm for collecting data on brain activity and the speech task necessary for this, creating an experimental setup, conducting experiments, developing decoding algorithms and studying their properties.

Panels a) and b) of 32 show MRI scans of two patients with electrode coordinates that were extracted from the corresponding post-implantation CT images. In addition, panels c) and d) show the mutual information profiles between neural activity and speech envelope for each of the electrodes. As a control curve displaying the values of mutual information under the null hypothesis of the absence of a relationship between

neuronal activity measured by the electrode and speech, the curve of mutual information between acoustic data is shown reversed in time with respect to the neural activity data. In these graphs, we see a distinct subset of electrodes with relatively high values of mutual information. Also, MRI images provide information on the result of mapping the speech cortex using electrical stimulation applied to the pairs of electrodes. As we can see, the location of pairs of electrodes, whose stimulation leads to violation of the speech function, correlates well with the position of the electrodes with high mutual information.

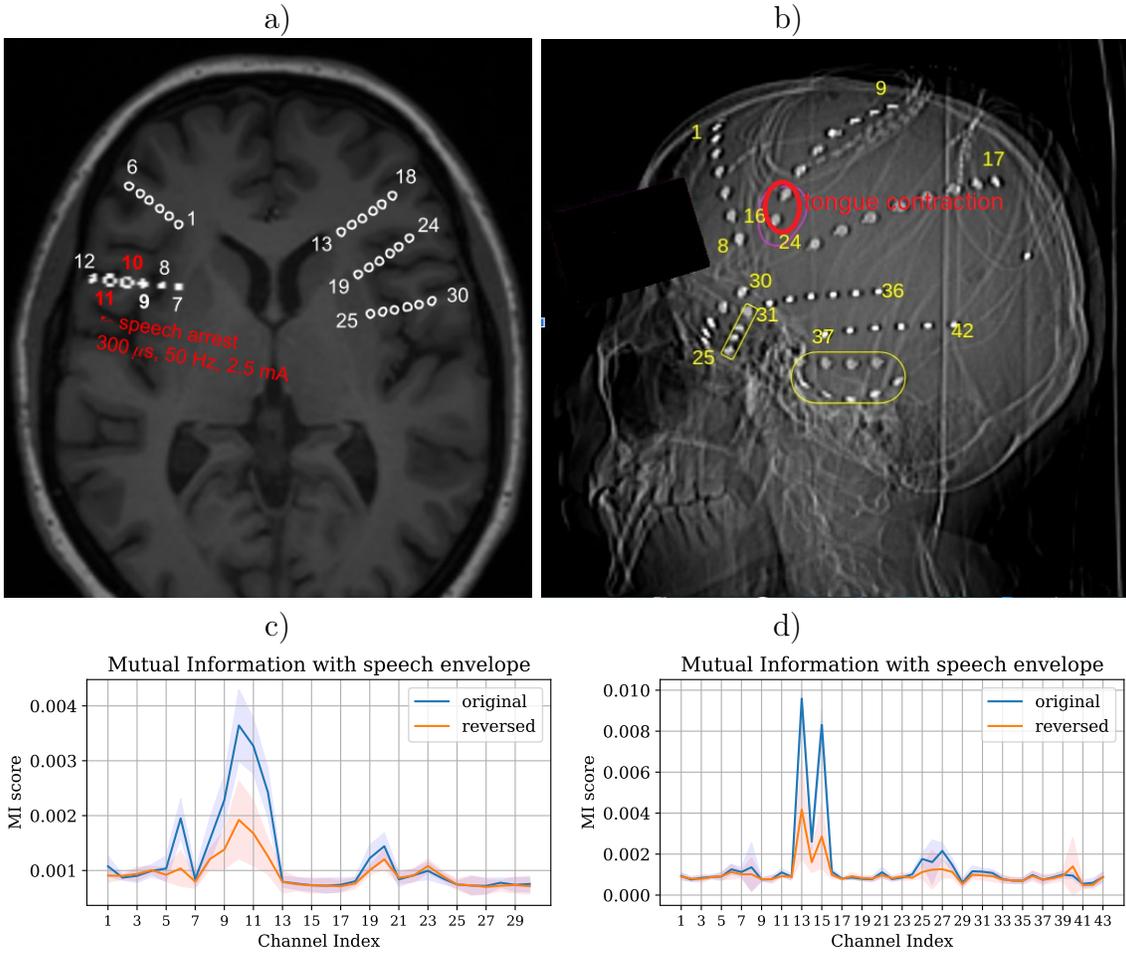


Рис. 32: Figure 1: a) stereo-EEG contacts extracted from the post-implantation CT image of the first patient superimposed on the corresponding MRI. Bipolar electrical stimulation of a pair of electrodes 10-11 (300  $\mu$ s, 2.5 mA, 50 Hz) resulted in a reproducible speech arrest. b) CT scan of the second patient who was implanted with nine 8-pin ECoG strips covering the frontal and lower temporal lobes on both sides. Bipolar electrical stimulation applied to electrodes 15-16 caused involuntary retraction of the tongue. c) Patient 1, profile of mutual information between the speech and stereo EEG envelopes in the gamma range (60 Hz-100 Hz). d) Patient 2, profile of mutual information between the speech envelope and the ECoG envelope in the gamma range (60 Hz-100 Hz)

### 5.2.8 Synchronous decoding

Our goal is to decode a finite number of words from the immediately preceding pieces of neural activity data. The direct approach, based on the neural network mapping from neural activity to the word number in the dictionary, requires the collection of a very large amount of training data. Instead, we developed our decoding solution based on the idea described in [Makin et al. \(2020\)](#), where a compact internal speech representation (ISR), similar to that used in vocoders, acts as a learning regularizer in the network, realizing the decoding of neural activity directly into words. In our approach, instead of using the ISR as a regularizer, we treat it as an intermediate decoding target. In other words, we first use our compact and interpretable [Petrosyan et al. \(2021a\)](#) architecture extended with LSTM to decode vector ISR time sequence from measurements of brain activity using sEEG or ECoG. We used the average correlation coefficient between actual and decoded ISRs as a quality criterion. Next, we fix the weights of our interpretable compact decoder and train a simple convolutional neural network to recognize discrete words based on the internal representations that have been generated in the ISR decoder. After training, our two-stage architecture works as a single network on minimal neural activity data and produces a discrete classification of individual words or the silent state. To train the network for the word classification task, we manually extracted segments of neural activity data corresponding to each word. We've also added a "silent" class, which corresponds to the intervals of silence between spoken words. When calculating accuracy scores, we randomly discarded some of the examples from the "silence" class to ensure class balance.

In this paradigm, the decoding of the Internal Speech Representation (ISR) vector time series from neural data was performed asynchronously, i.e. based on the sliding window aligned with each point in time and stretching into the past 1000 ms prior to the point in time for which the ISR vector was evaluated. The task of decoding individual words was then performed synchronously, i.e., based on the decoded speech representations cut out around the time moment each word was actually spoken.

### 5.2.9 Asynchronous decoding

We also experimented with a fully asynchronous approach, shown in [33](#). Unlike the synchronous mode, in the asynchronous approach our task is to predict the spoken word based on neural activity data preceding the spoken word (or silence interval) at *every* time  $t$ , so hypothetically this information can then be used to generate speech.

In the asynchronous mode, we first solve the problem of estimating the probability time profiles  $p_i(t)$  for each  $i$ -th word + silence class for each time  $t$  based on neural activity data  $[\mathbf{x}(tT), \dots, \mathbf{x}(t)]$  from the previous time window of length  $T$ . The next step is to smooth the resulting probability profiles and select a word (or silence) based on thresholding the smoothed probability profile  $\tilde{p}_i(t)$ . If  $\tilde{p}_i(t)$  reaches its maximum and exceeds the threshold, we make the appropriate decision and "pronounce" the  $i$ -th word. Importantly, we are assuming that this word cannot be spoken again until  $\tilde{p}_i(t)$  drops below, crosses the threshold, and peaks again. Our approach is described in more detail in [Petrosyan et al. \(2021b\)](#), including the technique for calculating precision-recall (PR) curves that characterize the quality of the algorithm.

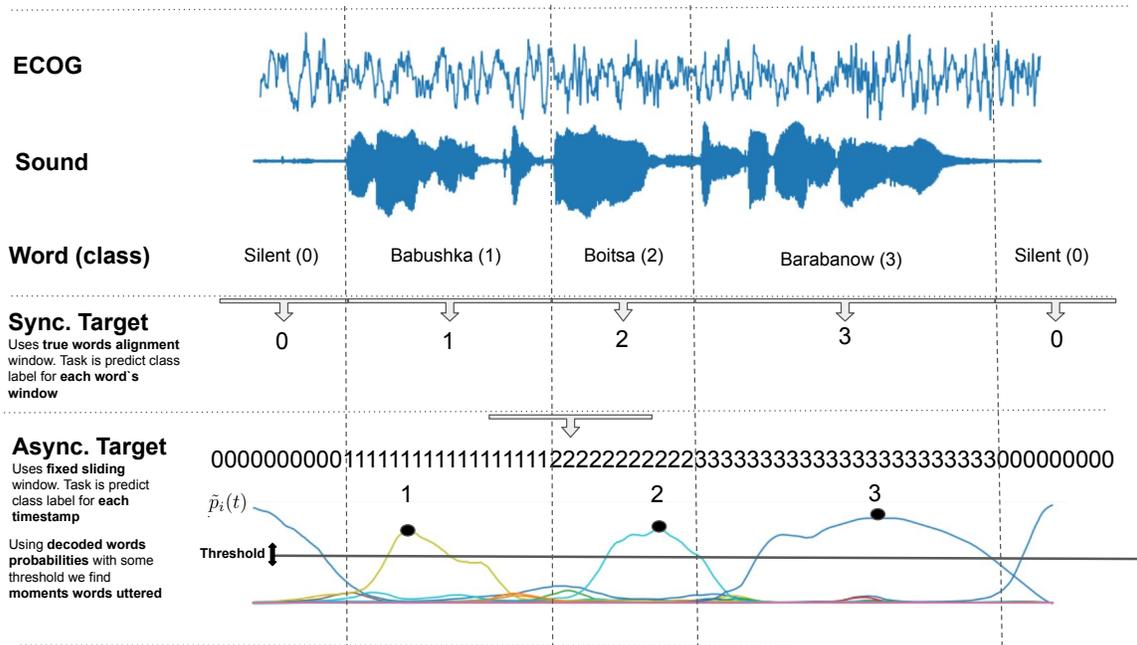


Рис. 33: Illustration of synchronous and asynchronous modes of operation. Unlike the synchronous mode, in the asynchronous mode our task is to predict the spoken word based on neural activity data for *each* time  $t$ .

### 5.3 Network architecture

For ISR decoding, we used the compact and interpretable convolutional network architecture previously developed for motor BCI [Petrosyan et al. \(2021a\)](#) and supplemented it with one bidirectional LSTM layer with 30 hidden units for compact modeling of temporal patterns. The LSTM layer is followed by a fully connected layer with the number of output neurons equal to one element of the ISR vector, the time profile of which we are trying to reconstruct based on the data of neuronal activity, see figure 34.

Note to the reader that, unlike [Angrick et al. \(2020\)](#), we do not specify feature extraction parameters in advance and let our architecture infer them during training, guided by the optimization of the average Pearson correlation coefficient between the original and decoded ISR time series.

### 5.4 Results

Figure 35 for two patients shows an example of true and reconstructed from the data of neuronal activity using our compact neural network logarithmic mel-spectrograms of the acoustic signal.

We experimented with various internal speech representations (ISRs). Figure 36 shows a comparison of the decoding accuracy achieved for different ISRs: LPC - linear prediction coefficients, LSF - spectral frequency lines, RC - reflection coefficients, LAR - log area ratio, LMSCs - log mel-spectrograms, MFCC - mel-cepstral coefficients. The left panel corresponds to the correlation coefficients between real and decoded ISR temporal profiles calculated for the entire temporal range of the test data segment. In

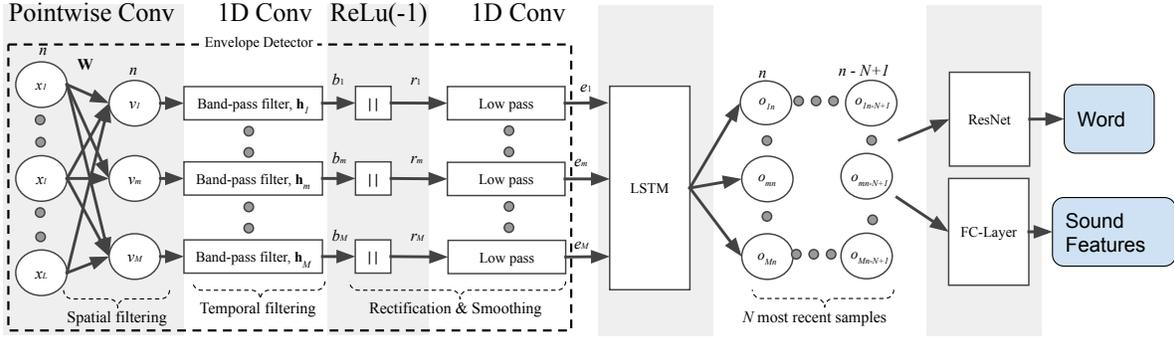


Рис. 34: Speech decoding architecture used in [Petrosyan et al. \(2021b\)](#). It was based on the structure of the interpreted network presented in [Petrosyan et al. \(2021a\)](#), which we adapted for the task of speech decoding by adding a bidirectional LSTM layer. Next, to solve the problem of decoding discrete words, we used a separate 2-D convolutional network, to the input of which we fed the vector of internal representation that arose in the LSTM layer during training for the ISR recovery problem.

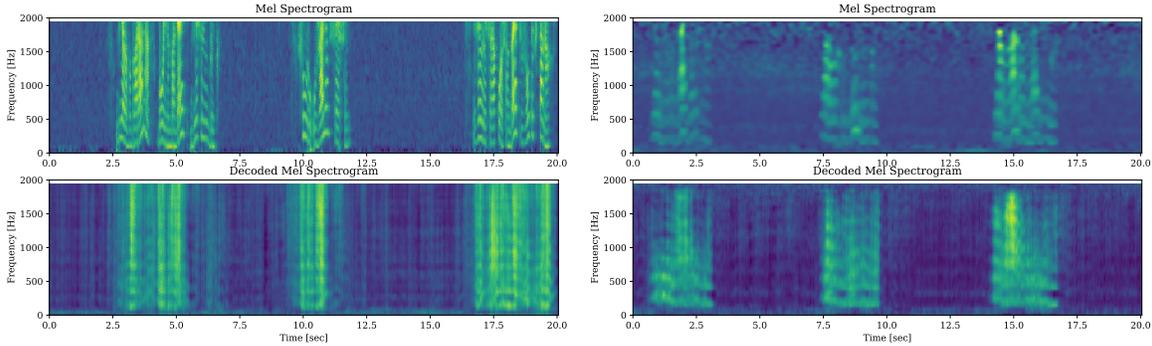


Рис. 35: An example of the true and reconstructed sequence of the ISR vector. Here, the vector of the logarithm of the mel-spectral coefficients was used as the ISR.

the right panel, the correlation coefficient is calculated only for time intervals in which real speech was present, that is, without taking into account silence intervals.

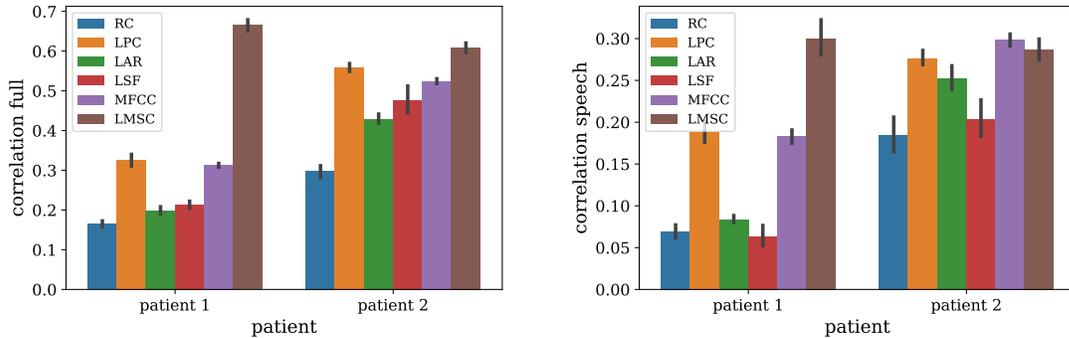


Рис. 36: Comparison of recovery accuracy of various ISRs. The left panel - the correlation coefficient calculated over the entire test time interval, on the right - only during the pronunciation of words.

As can be seen from 36, our compact architecture, using only 6 channels from a single stereo EEG wire, achieved almost 70% mean correlation over  $M = 40$  LMSC in patient 1 and 60% in patient 1 2 using the signal from one 8-channel ECoG strip. These accuracy values for decoding the internal speech representation are comparable to those reported in Angrick et al. (2019), which used a significantly larger number of data channels collected by multiple stereo EEG spokes.

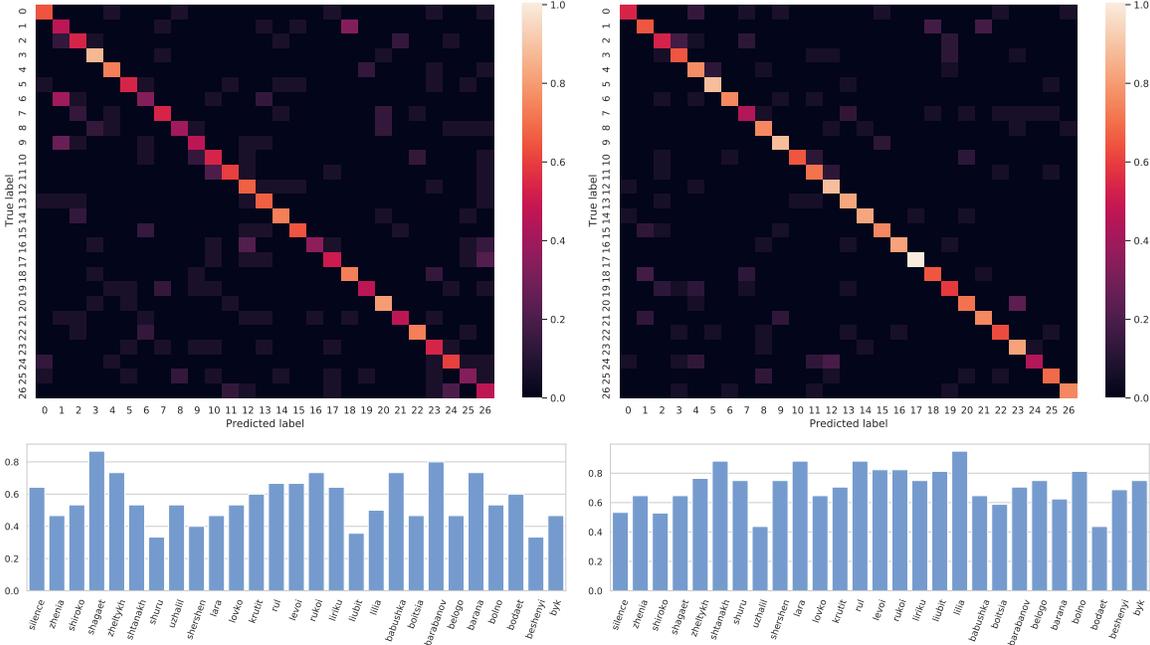


Рис. 37: Confusion matrix of classified words for patient 1 and patient 2. Words list: 0. silence, 1. zhenia, 2. shiroko, 3. shagaet, 4. zhelytykh, 5. shtanakh, 6. shuru, 7. uzhalil, 8. shershen, 9. lara, 10. lovko, 11. krutit, 12. rul, 13. levoi, 14. rukoi, 15. liriku, 16. liubit, 17. lilia, 18. babushka, 19. boitsia, 20. barabanov, 21. belogo, 22. barana, 23. bolno, 24. bodaet, 25. beshenyi, 26. byk. In the bottom we show the individual word decoding accuracy values, corresponding to the diagonal of the confusion matrix

The result of applying our weight interpretation procedure to each of the three branches of the neural network is shown in 38. From the top row of profiles corresponding to the first branch of the decision rule for Patient 1, it can be seen that the power spectral density (PSD) profile occupies a high frequency range of 100-200 Hz, and the corresponding spatial profile is limited to only one channel with an index of 11. At the same time, the second a branch with a much more scattered spatial structure, the spatial representation of which occupies channels 8-11, is characterized by PSD localized in the middle frequency range of 10-40 Hz. The inverse space-frequency relationship, which characterizes neuronal activity and distinguishes it from electromuscular artifacts, is also very well expressed in the second patient. Moving from top to bottom on the graphs of patient 2, we observe a gradual increase in spatial dispersion with the migration of the PSD frequency range from higher frequencies to lower ones.

In combination with knowledge of the subject area Buzsaki (2006); Buzsáki et al. (2012); Brunner et al. (2009); Volkova et al. (2019b), emphasizing the mutual spatiotemporal relationships in the observed profiles of cortical activity and

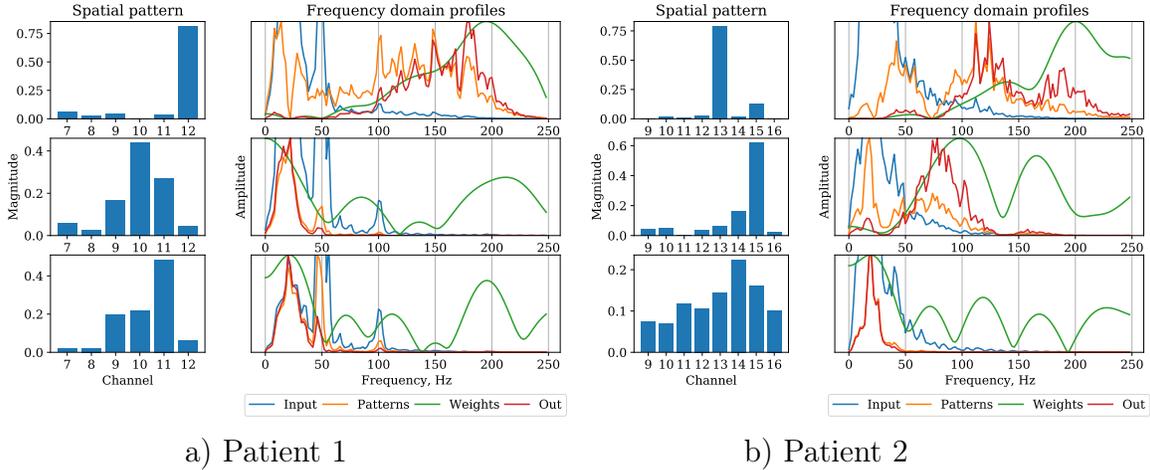


Рис. 38: A theoretically sound interpretation of the weights applies to the most significant branches of the architecture in 29.

phenomenological observations [Eliseyev and Aksenova \(2014\)](#) about the properties of electromuscular activity and its representation in the cerebral cortex, the observed combinations of spatial and frequency profiles allow us to infer the neural origin of the sources of information to which our decoder tuned in during the learning process. Microphone effect analysis, detailed in [Petrosyan et al. \(2021b\)](#), also virtually eliminates the chances that decoding is based on acoustic signal leakage into neural data channels.

Next, we compared the decoding accuracy achievable with our compact architecture with that provided by more sophisticated neural network solutions used in a number of recent studies on decoding speech from invasive brain electrical activity data. Panel a) of Figure 39 shows the corresponding decoding accuracy values. The accuracy of a word classifier based on features extracted using our compact architecture (Envelope Detector net, ED-Net) in combination with the LSTM layer was reliably higher than the accuracy obtained when ResNet was used instead of ED-Net, known for its high performance achievable for by minimizing the attenuation of the gradient by organizing additional end-to-end connections. One reason for this state of affairs may be due to the relatively small amount of data available for training our networks. For each patient, we have at our disposal about 1 hour of recording containing about 100 - 150 repetitions of each of the 26 words of which we use 70% for training. The compactness of our architecture and the correspondingly small number of adjustable parameters, combined with high decoding accuracy, make it an optimal candidate for such systems.

We also compared the decoding accuracies achievable using different internal speech representations (ISR) (panel b), in causal, anticausal, or non-causal mode (panel c), as well as using different subsets of electrodes (panel d). A detailed analysis of the results is given in the article [Petrosyan et al. \(2021b\)](#).

## 6 Conclusion

The following papers reflect main results of this dissertatio: [Kuznetsova et al. \(2021\)](#); [Greenblatt et al. \(2005c\)](#); [Ossadtchi et al. \(2018\)](#); [Greenblatt et al. \(2012\)](#); [Kozunov and Ossadtchi \(2015\)](#); [Kleeva et al. \(2022\)](#); [Ossadtchi et al. \(2004a, 2005\)](#); [Kuznetsova](#)

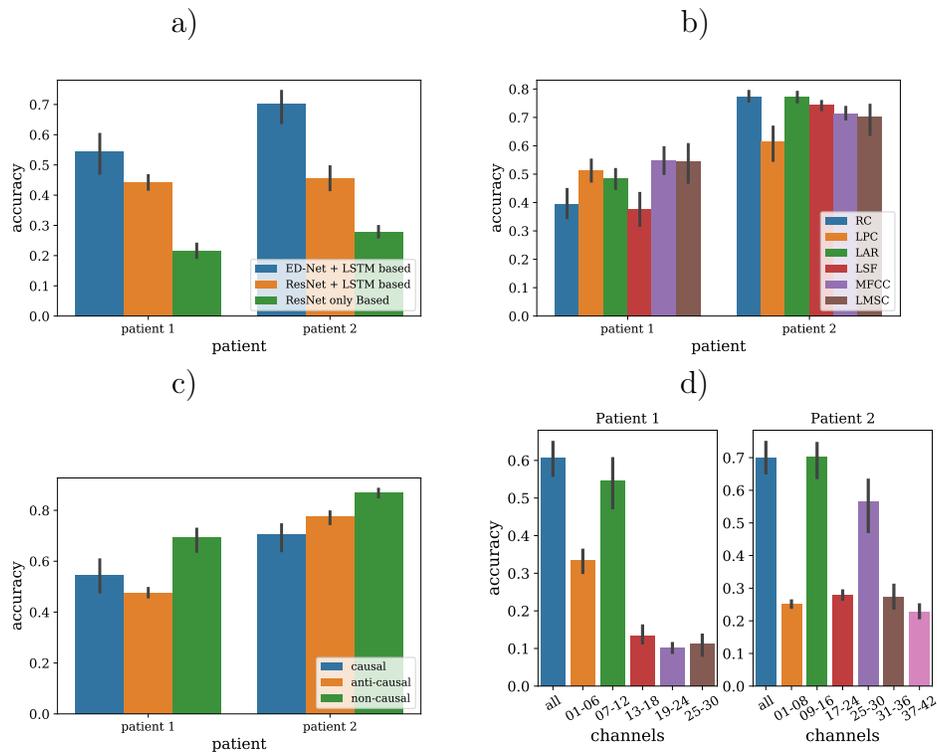


Рис. 39: Comparative analysis. a) Comparison of various neural network models b) Comparison of various possible intermediate audio representations, LPC - autoregressive prediction coefficients, LSF - linear spectral frequencies, RC - reflection coefficients, LAR - logarithmic area ratios, LMSC - logarithmic chalk spectrogram coefficients, MFCC - cepstral coefficients mel-frequencies c) Comparison of different possible delays d) Comparison of decoding quality for different channel subsets.

and Ossadtchi (2022); Ossadtchi et al. (2010); Smetanin et al. (2020a); Belinskaia et al. (2020a); Ossadtchi et al. (2017a); Petrosyan et al. (2021a,b). The papers Kuznetsova et al. (2021); Greenblatt et al. (2005c); Ossadtchi et al. (2018); Greenblatt et al. (2012); Kozunov and Ossadtchi (2015) describe new methods for solving the EEG and MEG inverse problem, including those aimed at finding functional relationships between areas of the cerebral cortex. The articles Kleeva et al. (2022); Kuznetsova and Ossadtchi (2022) introduce mathematical methods for detecting the epileptogenic zone using MEG data and extend the earlier developments of the applicant Ossadtchi et al. (2010, 2005, 2004a). The next cycle of works is devoted to methods of EEG processing in real-time systems. The publications Ossadtchi et al. (2017a); Smetanin et al. (2020a); Belinskaia et al. (2020a) discuss the phenomenological and algorithmic aspects of neurofeedback technology and suggest methods for low-latency real-time estimation of brain's rhythmic activity and Petrosyan et al. (2021a,b) are devoted to the description of a new compact architecture for decoding multichannel electrophysiological data and methods for interpreting the weight coefficients, with the help of which it is possible to invser novel knowledge and censor the learned decision rule.

## The main results of the work submitted for defence

- Projection operation in the  $R^N \times R^N$  product-space of magneto- and electro-encephalographic (MEEG) data covariance matrices for beamforming in the environment with correlated sources
- Projection operation in the  $R^N \times R^N$  product-space of magneto- and electro-encephalographic (MEEG) data covariance matrices for detecting functionally coupled sources characterized by a small mutual phase delay
- An iterative Bayesian approach for the improved efficiency of solving the MEG inverse problem at the group level through by exploiting inter-individual similarities and differences in the spatial characteristics of neuronal activity
- Biomimetic convolutional method for detecting interictal events in multichannel recordings of brain electrical activity
- A technique for solving the MEG inverse problem for interictal activity using the cortical traveling wave model and its application for epileptogenic zone localization
- Low-latency methods for assessing parameters of brain's rhythmic activity in real time as the basis for instantaneous neurofeedback technology
- Neural network architecture for decoding brain activity signals with applications to motor and speech neurointerfaces. The theoretically justified method for interpreting the neural network's weights for decision rule censoring and discovery of geometric and dynamic properties of the pivotal to the task neuronal populations.

## References

- Alexander, D. M., Trengove, C., and van Leeuwen, C. (2015). Donders is dead: cortical traveling waves and the limits of mental chronometry in cognitive neuroscience. *Cognitive Processing*, 16(4):365–375.
- Angrick, M., Herff, C., Mugler, E., Tate, M. C., Slutzky, M. W., Krusienski, D. J., and Schultz, T. (2019). Speech synthesis from ecog using densely connected 3d convolutional neural networks. *Journal of neural engineering*, 16(3):036019.
- Angrick, M., Ottenhoff, M., Diener, L., Ivucic, D., Ivucic, G., Goulis, S., Saal, J., Colon, A. J., Wagner, L., Krusienski, D. J., et al. (2020). Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. *bioRxiv*.
- Belinskaia, A., Smetanin, N., Lebedev, M., and Ossadtchi, A. (2020a). Short-delay neurofeedback facilitates training of the parietal alpha rhythm. *Journal of Neural Engineering*, 17(6):066012.
- Belinskaia, A., Smetanin, N., Lebedev, M., and Ossadtchi, A. (2020b). Short-delay neurofeedback facilitates training of the parietal alpha rhythm. *Journal of Neural Engineering*, 17(6):066012.
- Bergmann, T. O., Karabanov, A., Hartwigsen, G., Thielscher, A., and Siebner, H. R. (2016). Combining non-invasive transcranial brain stimulation with neuroimaging and electrophysiology: Current approaches and future perspectives. *NeuroImage*, 140:4–19.
- Borgiotti, G. and Kaplan, L. (1979). Superresolution of uncorrelated interference sources by using adaptive array techniques. *IEEE Transactions on Antennas and Propagation*, 27(6):842–845.
- Brunner, P., Ritaccio, A. L., Lynch, T. M., Emrich, J. F., Wilson, J. A., Williams, J. C., Aarnoutse, E. J., Ramsey, N. F., Leuthardt, E. C., Bischof, H., et al. (2009). A practical procedure for real-time functional mapping of eloquent cortex using electrocorticographic signals in humans. *Epilepsy & Behavior*, 15(3):278–286.
- Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford University Press.
- Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, 13(6):407–420.
- Chen, L. L., Madhavan, R., Rapoport, B. I., and Anderson, W. S. (2011). Real-time brain oscillation detection and phase-locked stimulation using autoregressive spectral estimation and time-series forward prediction. *IEEE transactions on Biomedical Engineering*, 60(3):753–762.
- Dalal, S., Sekihara, K., and Nagarajan, S. (2006). Modified beamformers for coherent source region suppression. *IEEE Transactions on Biomedical Engineering*, 53(7):1357–1363.

- De Oliveira, P. G., Queiroz, C., and Da Silva, F. L. (1983). Spike detection based on a pattern recognition approach using a microcomputer. *Electroencephalography and clinical neurophysiology*, 56(1):97–103.
- Diamond, J. M., Diamond, B. E., Trotta, M. S., Dembny, K., Inati, S. K., and Zaghoul, K. A. (2021). Travelling waves reveal a dynamic seizure source in human focal epilepsy. *Brain*, 144(6):1751–1763.
- Dingle, A. A., Jones, R. D., Carroll, G. J., and Fright, W. R. (1993). A multistage system to detect epileptiform activity in the eeg. *IEEE Transactions on Biomedical Engineering*, 40(12):1260–1268.
- Duin, R. P. and Pękalska, E. (2010). Non-euclidean dissimilarities: Causes and informativeness. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 324–333. Springer.
- Eliseyev, A. and Aksenova, T. (2014). Stable and artifact-resistant decoding of 3d hand trajectories from ecog signals using the generalized additive model. *Journal of neural engineering*, 11(6):066005.
- Elmarakeby, H. A., Hwang, J., Arafeh, R., Crowdis, J., Gang, S., Liu, D., AlDubayan, S. H., Salari, K., Kregel, S., Richter, C., et al. (2021). Biologically informed deep neural network for prostate cancer discovery. *Nature*, 598(7880):348–352.
- Engel A.K., Fries P., S. W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.*, 2001(2):704–716.
- Ewald, A., Avarvand, F. S., and Nolte, G. (2014). Wedge music: A novel approach to examine experimental differences of brain source connectivity patterns from eeg/meg data. *NeuroImage*, 101:610–624.
- Faure, C. (1985). Attributed strings for recognition of epileptic transients in eeg. *International journal of bio-medical computing*, 16(3-4):217–229.
- Fries, P. (2015). Rhythms for cognition: Communication through coherence. *Neuron*, 88(1):220–235.
- Friston, K. (2002). Functional integration and inference in the brain. *Progress in neurobiology*, 68(2):113–143.
- Glover, J. R., Ktonas, P. Y., Raghavan, N., Urunuela, J. M., Velamuri, S. S., and Reilly, E. L. (1986). A multichannel signal processor for the detection of epileptogenic sharp transients in the eeg. *IEEE transactions on biomedical engineering*, (12):1121–1128.
- Gotman, J. and Gloor, P. (1976). Automatic recognition and quantification of interictal epileptic activity in the human scalp eeg. *Electroencephalography and clinical neurophysiology*, 41(5):513–529.
- Gotman, J. and Wang, L. (1991). State-dependent spike detection: concepts and preliminary results. *Electroencephalography and clinical Neurophysiology*, 79(1):11–19.

- Gotman, J. and Wang, L.-Y. (1992). State dependent spike detection: validation. *Electroencephalography and clinical neurophysiology*, 83(1):12–18.
- Greenblatt, R., Ossadtchi, A., and Pflieger, M. (2005a). Local linear estimators for the bioelectromagnetic inverse problem. *IEEE Trans Signal Proc.*
- Greenblatt, R., Ossadtchi, A., and Pflieger, M. (2005b). Local linear estimators for the bioelectromagnetic inverse problem. *IEEE Transactions on Biomedical Engineering*, 53(9):3403–3412.
- Greenblatt, R. E., Ossadtchi, A., and Pflieger, M. E. (2005c). Local linear estimators for the bioelectromagnetic inverse problem. *IEEE transactions on signal processing*, 53(9):3403–3412.
- Greenblatt, R. E., Pflieger, M., and Ossadtchi, A. (2012). Connectivity measures applied to human brain electrophysiological data. *Journal of neuroscience methods*, 207(1):1–16.
- Halford, J. J., Schalkoff, R. J., Zhou, J., Benbadis, S. R., Tatum, W. O., Turner, R. P., Sinha, S. R., Fountain, N. B., Arain, A., Pritchard, P. B., et al. (2013). Standardized database development for eeg epileptiform transient detection: Eegnet scoring system and machine learning analysis. *Journal of neuroscience methods*, 212(2):308–316.
- Hamalainen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews on modern physics*, 65(2).
- Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, 87:96–110.
- Hincapié, A. S., Kujala, J., Mattout, J., Daligault, S., Delpuech, C., Mery, D., Cosmelli, D., and Jerbi, K. (2016). MEG connectivity and power detections with minimum norm estimates require different regularization parameters. *Computational Intelligence and Neuroscience*, 2016(March).
- Hostetler, W. E., Doller, H. J., and Homan, R. W. (1992). Assessment of a computer program to detect epileptiform spikes. *Electroencephalography and clinical neurophysiology*, 83(1):1–11.
- Jaiswal, A., Nenonen, J., Stenroos, M., Gramfort, A., Dalal, S. S., Westner, B. U., Litvak, V., Mosher, J. C., Schoffelen, J.-M., Witton, C., et al. (2020). Comparison of beamformer implementations for meg source localization. *NeuroImage*, 216:116797.
- Keshri, A. K., Sinha, R. K., Singh, A., and Das, B. N. (2011). Dfspike: A new computational proposition for efficient recognition of epileptic spike in eeg. *Computers in biology and medicine*, 41(7):559–564.
- Kleeva, D., Soghoian, G., Komoltsev, I., Sinkin, M., and Ossadtchi, A. (2022). Fast parametric curve matching (fpcm) for automatic spike detection. *Journal of Neural Engineering*, 19(3):036003.

- Kozunov, V. V. and Ossadtchi, A. (2015). Gala: group analysis leads to accuracy, a novel approach for solving the inverse problem in exploratory analysis of group meg recordings. *Frontiers in Neuroscience*, 9:107.
- Kuznetsova, A., Nurislamova, Y., and Ossadtchi, A. (2021). Modified covariance beamformer for solving meg inverse problem in the environment with correlated sources. *Neuroimage*, 228:117677.
- Kuznetsova, A. and Ossadtchi, A. (2022). Анализ локальной динамики распространения межприступных разрядов с помощью модели бегущих волн. *Журнал высшей нервной деятельности им. И.П. Павлова*, 1(3):370–386.
- Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4):194–208. cited By (since 1996)852.
- Larson, E., Maddox, R. K., and Lee, A. K. (2014). Improving spatial localization in meg inverse imaging by leveraging intersubject anatomical differences. *Frontiers in neuroscience*, 8:330.
- Lebedev, M. A. and Ossadtchi, A. (2018). Bidirectional neural interfaces. In *Brain–Computer Interfaces Handbook*, pages 701–720. CRC Press.
- Liu, Y.-C., Lin, C.-C. K., Tsai, J.-J., and Sun, Y.-N. (2013). Model-based spike detection of epileptic eeg data. *Sensors*, 13(9):12536–12547.
- Makin, J. G., Moses, D. A., and Chang, E. F. (2020). Machine translation of cortical activity to text with an encoder–decoder framework. *Nature Neuroscience*, 23(4):575–582.
- McIntosh, J. R. and Sajda, P. (2020). Estimation of phase in EEG rhythms for real-time applications. *Journal of Neural Engineering*, 17(3):034002.
- Mosher, J. and Leahy, R. (1999a). Source localization using recursively applied and projected (RAP) MUSIC. *IEEE Transactions on Signal Processing*, 47(2):332–340.
- Mosher, J. C. and Leahy, R. M. (1999b). Source localization using recursively applied and projected (rap) music. *IEEE Transactions on signal processing*, 47(2):332–340.
- Müller, K.-R., Krauledat, M., Dornhege, G., Curio, G., and Blankertz, B. (2004). Machine learning techniques for brain-computer interfaces. *Biomed. Tech*, 49(1):11–22.
- Nasiotis, K., Clavagnier, S., Baillet, S., and Pack, C. C. (2017). High-resolution retinotopic maps estimated with magnetoencephalography. *NeuroImage*, 145:107–117.
- Nishida, S., Nakamura, M., Ikeda, A., and Shibasaki, H. (1999). Signal separation of background eeg and spike by using morphological filter. *Medical engineering & physics*, 21(9):601–608.

- Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, 115(10):2292–2307.
- Ossadtchi, A., Altukhov, D., and Jerbi, K. (2018). Phase shift invariant imaging of coherent sources (PSIICOS) from MEG data. *NeuroImage*, 183:950–971.
- Ossadtchi, A., Baillet, S., Mosher, J., Thyerlei, D., Sutherling, W., and Leahy, R. (2004a). Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering. *Clinical Neurophysiology*, 115(3):508–522.
- Ossadtchi, A., Baillet, S., Mosher, J., Thyerlei, D., Sutherling, W., and Leahy, R. (2004b). Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering. *Clinical Neurophysiology*, 115(3):508–522.
- Ossadtchi, A., Greenblatt, R., Towle, V., Kohrman, M., and Kamada, K. (2010). Inferring spatiotemporal network patterns from intracranial eeg data. *Clinical Neurophysiology*, 121(6):823–835.
- Ossadtchi, A., Mosher, J. C., Sutherling, W. W., Greenblatt, R. E., and Leahy, R. M. (2005). Hidden markov modelling of spike propagation from interictal MEG data. *Physics in Medicine and Biology*, 50(14):3447–3469.
- Ossadtchi, A., Shamaeva, T., Okorokova, E., Moiseeva, V., and Lebedev, M. (2017a). Neurofeedback learning modifies the incidence rate of alpha spindles, but not their duration and amplitude. *Scientific Reports*.
- Ossadtchi, A., Shamaeva, T., Okorokova, E., Moiseeva, V., and Lebedev, M. A. (2017b). Neurofeedback learning modifies the incidence rate of alpha spindles, but not their duration and amplitude. *Scientific reports*, 7(1):1–12.
- O’Neill, G. C., Barratt, E. L., Hunt, B. A., Tewarie, P. K., and Brookes, M. J. (2015). Measuring electrophysiological connectivity by power envelope correlation: a technical review on meg methods. *Physics in Medicine & Biology*, 60(21):R271.
- Papanicolaou, A. C. (1998). *Fundamentals of functional brain imaging: A guide to the methods and their applications to psychology and behavioral neuroscience*. CRC Press.
- Petrosyan, A., Lebedev, M., and Ossadtchi, A. (2020a). Decoding neural signals with a compact and interpretable convolutional neural network. In *International Conference on Neuroinformatics*, pages 420–428. Springer.
- Petrosyan, A., Lebedev, M., and Ossadtchi, A. (2020b). Linear systems theoretic approach to interpretation of spatial and temporal weights in compact cnns: Monte-carlo study. In *Biologically Inspired Cognitive Architectures Meeting*, pages 365–370. Springer.

- Petrosyan, A., Sinkin, M., Lebedev, M., and Ossadtchi, A. (2021a). Decoding and interpreting cortical signals with a compact convolutional neural network. *Journal of Neural Engineering*, 18(2):026019.
- Petrosyan, A., Voskoboynikov, A., and Ossadtchi, A. (2021b). Compact and interpretable architecture for speech decoding from stereotactic eeg. In *2021 Third International Conference Neurotechnologies and Neurointerfaces (CNN)*, pages 79–82. IEEE.
- Pikovsky, A., M., R., , Kurths, and J. (2001). *Synchronization. A Universal Concept in Nonlinear Sciences*. Cambridge University Press.
- Poldrack, R. A. (2018). *The new mind readers: What neuroimaging can and cannot reveal about our thoughts*. Princeton University Press.
- Popescu, M., Popescu, E.-A., Chan, T., Blunt, S., and Lewine, J. (2008). Spatio-temporal reconstruction of bilateral auditory steady-state responses using MEG beamformers. *IEEE Transactions on biomedical engineering*, 55(3):1092–1102.
- Rajagovindan, R. and Ding, M. (2008). Decomposing neural synchrony: Toward an explanation for near-zero phase-lag in cortical oscillatory networks. *PLoS One*, 3(11):e3649.
- Rizzolatti, G., Fabbri-Destro, M., Caruana, F., and Avanzini, P. (2018). System neuroscience: Past, present, and future. *CNS Neuroscience & Therapeutics*, 24(8):685–693.
- Roelfsema, P., Engel, A., Knig, P., and Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature*, 385(6612):157–161.
- Sabbagh, D., Ablin, P., Varoquaux, G., Gramfort, A., and Engemann, D. A. (2019). Manifold-regression to predict from meg/eeg brain signals without source modeling. *Advances in Neural Information Processing Systems*, 32.
- Sabbagh, D., Ablin, P., Varoquaux, G., Gramfort, A., and Engemann, D. A. (2020). Predictive regression modeling with meg/eeg: from source power to signals and cognitive states. *NeuroImage*, 222:116893.
- Schaworonkow, N., Caldana Gordon, P., Belardinelli, P., Ziemann, U., Bergmann, T. O., and Zrenner, C. (2018).  $\mu$ -rhythm extracted with personalized eeg filters correlates with corticospinal excitability in real-time phase-triggered eeg-tms. *Frontiers in Neuroscience*, 12:954.
- Scheuer, M. L., Bagic, A., and Wilson, S. B. (2017). Spike detection: Inter-reader agreement and a statistical turing test on a large data set. *Clinical Neurophysiology*, 128(1):243–250.
- Schuetze, S. M. (1983). The discovery of the action potential. *Trends in Neurosciences*, 6:164–168.

- Schuster, H. G. and Wagner, P. (1989). Mutual entrainment of two limit cycle oscillators with time delayed coupling. *Progress of Theoretical Physics*, 81(5).
- Sekihara, K. and Nagarajan, S. (2008). *Adaptive Spatial Filters for Electromagnetic Brain Imaging*. Springer Berlin Heidelberg.
- Sekihara, K., Nagarajan, S., Poeppel, D., Marantz, A., and Miyashita, Y. (2001). Reconstructing spatio-temporal activities of neural sources using an MEG vector beamformer technique. *IEEE Transactions on Biomedical Engineering*, 48(7):760–771.
- Shakeel, A., Tanaka, T., and Kitajo, K. (2020). Time-series prediction of the oscillatory phase of eeg signals using the least mean square algorithm-based ar model. *Applied Sciences*, 10(10).
- Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron*, 1999(24):49–65.
- Smetanin, N., Belinskaya, A., Lebedev, M., and Ossadtchi, A. (2020a). Digital filters for low-latency quantification of brain rhythms in real time. *Journal of Neural Engineering*, 17(4):046022.
- Smetanin, N., Belinskaya, A., Lebedev, M., and Ossadtchi, A. (2020b). Digital filters for low-latency quantification of brain rhythms in real time. *Journal of Neural Engineering*, 17(4):046022.
- Smetanin, N., Volkova, K., Zabodaev, S., Lebedev, M. A., and Ossadtchi, A. (2018a). NFBLab—a versatile software for neurofeedback and brain-computer interface research. *Frontiers in Neuroinformatics*, 12.
- Smetanin, N., Volkova, K., Zabodaev, S., Lebedev, M. A., and Ossadtchi, A. (2018b). Nfblab—a versatile software for neurofeedback and brain-computer interface research. *Frontiers in neuroinformatics*, 12:100.
- Smith, E. H., Liou, J.-y., Merricks, E. M., Davis, T., Thomson, K., Greger, B., House, P., Emerson, R. G., Goodman, R., McKhann, G. M., et al. (2022). Human interictal epileptiform discharges are bidirectional traveling waves echoing ictal discharges. *Elife*, 11:e73541.
- Staley, K. J. and Dudek, F. E. (2006). Interictal spikes and epileptogenesis. *Epilepsy Currents*, 6(6):199–202.
- Stam, C., Nolte, G., and Daffertshofer, A. (2007). Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum Brain Mapp.*, 28(11):1178–93.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1):267–288.

- Tomlinson, S. B., Bermudez, C., Conley, C., Brown, M. W., Porter, B. E., and Marsh, E. D. (2016). Spatiotemporal mapping of interictal spike propagation: A novel methodology applied to pediatric intracranial EEG recordings. *Frontiers in Neurology*, 7.
- Van Veen, B., Van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 44(9):867–880.
- Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2(4):229–239.
- Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F., and Pennartz, C. (2011). A improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage*, 55(4):1548–65.
- Volkova, K., Lebedev, M. A., Kaplan, A., and Ossadtchi, A. (2019a). Decoding movement from electrocorticographic activity: A review. *Frontiers in Neuroinformatics*, 13.
- Volkova, K., Lebedev, M. A., Kaplan, A., and Ossadtchi, A. (2019b). Decoding movement from electrocorticographic activity: A review. *Frontiers in neuroinformatics*, 13:74.
- Webber, W. R. S., Litt, B., Lesser, R. P., Fisher, R., and Bankman, I. (1993). Automatic eeg spike detection: what should the computer imitate? *Electroencephalography and clinical neurophysiology*, 87(6):364–373.