Mathematical Biology and Bioinformatics
2017. V. 12. № 1. P. 176–185. doi: 10.17537/2017.12.176

============================== DATA MINING ==============================

UDC: 612.825+004.75

Methods for magnetic encephalography data analysis in MathBrain cloud service

Rykunov S.D.[*1], Oplachko E.S.[1], Ustinin M.N.[1], Llinás R.R.[2]

[1]*Institute of Mathematical Problems of Biology RAS – the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, 142290, Russian Federation*
[2]*New York University, New York, NY, 10016, USA*

Abstract. MathBrain is a cloud-based application, distributed under Software as a Service model. This application provides access to several algorithms of the multichannel encephalography data analysis. Spectral methods include direct and inverse Fourier transforms and quantitative analysis. Statistical methods involve principal component analysis and independent component analysis. The field maps of elementary components can be used to solve the magnetic encephalography inverse problem and to display the result at the magnetic resonance image. The application is designed to be used in human brain studies without mathematical training.

Key words: *cloud computing, magnetic encephalography, spectral analysis, independent component analysis, inverse problem solution.*

INTRODUCTION

Magnetic encephalography (MEG) has become one of the foremost biological technologies addressing detailed analysis of human brain function. It is high-precision, non-invasive method, based on passive registration of electro-magnetic activity and generating large amount of detailed data. Magnetic encephalograph register magnetic field for several minutes, in hundreds of channels with sampling frequency up to several thousand Hertz. So, the problem of data analysis appears a pressing challenge in the MEG technique. Many approaches are used to solve various scientific and diagnostic problems of encephalography. Fourier analysis in many implementations can be called the oldest and the most popular of methods used for the brain data analysis [1–3]. Through the whole history of this method it was connected with difficulties of calculations, so the development of the Fast Fourier Transform (FFT) [4] dramatically advanced the application of the Fourier analysis in many fields, including brain research [5]. In the quantitative electric and magnetic encephalography studies, trying to study patterns of the Fourier coefficients, rough spectral analysis is combined with statistical analysis of coherence between channels or independent components [6–10]. Usually in applications of the Fourier analysis to brain studies the spectra are calculated in short (< 10 s) time windows, based on the well-known property of instability of the brain processes [11].

Recently the method of precise frequency-pattern analysis to decompose complex systems into functionally invariant entities was proposed [12, 13]. The method makes it possible to address general spectra to the partial spectra of stable functional entities and to restore their time series. The method is based on the complete utilization of the long-time series, while the multichannel nature of the data is also completely taken into account, making it possible to extract elementary sources of the brain activity. It was successively applied in alpha-rhythm studies [14] and in partial spectroscopy of the brain [15]. Mathematical foundations of this method were implemented in the MEGMRIAn software [16].

There are several academic software applications for the MEG data analysis, such as FieldTrip [17], EEGLAB [18], BrainStorm [19] and others. These software pieces were designed as MATLAB Toolboxes, so they cannot be used without installation of MATLAB. Also, due to high-complexity and large size of MEG data, to analyze such data one require having a computer with above the average computational power. The MathBrain cloud service [http://www.mathbrain.ru], proposed in this paper, is free from such limitations. It is based on free and open-source technologies, utilizes computational power of high-performance cluster and can be accessed from any place, using only a web-browser. It provides simple and intuitive access to precise frequency-pattern analysis and other MEG analysis methods. The MathBrain cloud application is distributed under Software as a Service model. It means, that the user has access to computational facilities and to the hardware, installed at these facilities [20].

The following methods of the MEG analysis are now available at MathBrain: direct and inverse Fourier transforms, principal component and independent component decompositions and quantitative analysis. Multichannel elementary oscillation, extracted in some or other technique, can be characterized by the map of magnetic field. From this map, the inverse problem solution in single dipole model can be obtained. Here we present the results of data processing, obtained at the experimental data set for control subject.

MathBrain software architecture

MathBrain consists of three main modules:

- front-end module;
- database and data-storage module;
- computational back-end module.

Each of these modules comes in its own Docker virtual container. Such implementation provides us with ability to easily migrate them between computational nodes and provides great scalability. User authentication, based on stored in database login-password pair is implemented. To provide security all passwords are stored as double MD5 hashes.

Front-end module includes the web-interface written in HTML and JavaScript languages, on server side there are Apache web-server, JSON-RPC sender and PHP scripts. It provides user with tools to create new data analysis task, review results and manage user’s data-files (e.g. rename, delete and change description). Web-interface utilizes power of Mpld3 library to provide user with rich interactive visualization tools, such as dynamically scaling plots and plot data pickers. JSON-RPC sender is used to provide communication between front-end and computational back-end. Examples of web-interface are shown in Fig. 1.

Database and data-storage module consists of file storage and MySQL database. Database is used to store file descriptions, history of analysis tasks and its parameters, file ownership and sharing settings. At the moment, MEG time-series recordings are stored in HDF5-based MATLAB file format, data structure (names of fields and MEG device description format) is close to FieldTrip [17] file format. Processing results are stored in MATLAB file format.

Computational back-end module consists of JSON-RPC listener/task-dispatcher script, SLURM resource manager and computational programs. RPC listener is a Flask application, responsible for communication between MathBrain’s front-end and SLURM scheduler, it is written in Python and uses RADICAL-SAGA library. It has the following methods: computational task creation, task status check and task cancellation. Usage of RADICAL-SAGA library allows us to use local SLURM-cluster as well as external cloud computational services, such Amazon EC. Computational programs are written in Python, using Scipy [21], Numpy [21], scikit-learn [22] and FFTW3 [23] libraries.

Mathematical biology and bioinformatics. 2017. V. 12. № 1. doi: 10.17537/2017.12.176

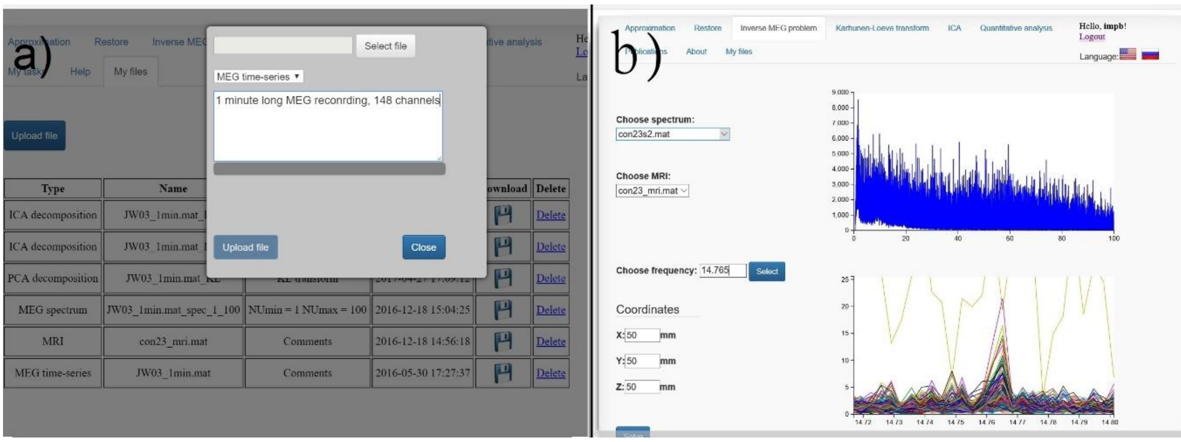

Fig. 1. Web-interface examples: **a) –** file upload form; **b) –** parameters for MEG inverse problem.

Data analysis workflow performs the following steps.

1. User logs in MathBrain service, using his login and password. After login, “my files” tab is showed, on which one can review existing files and upload new.
2. After clicking on “Upload new file” button, pop-up form with file selection dialog, type selection and comments text-box is shown. User selects file on his hard drive, selects proper type, enters comments and presses “Upload file” button. File starts to upload to server. After transfer is complete, pop-up form closes. Uploaded file receives unique name and ID, then it’s ID, original name, type, comments, and file system path are added to database. From this moment, uploaded file is available for analysis.
3. From tabs at top side of the page user selects analysis method. On corresponding tab user selects file to analyze and enters analysis parameters. After clicking on “Process” button analysis process starts. It consists of the following steps:
   - RPC-sender transmits processing parameters and file ID to computational back-end;
   - RPC-listener creates and submits computational task to the queue, adds task record to the database;
   - input file is transferred to assigned computational node, and task is executed, output file is transferred to file storage, database entry for output file is added;
   - results are available for viewing and downloading at “my tasks” tab.

Experimental data description

The complete set of the experimental data for brain studies consists of the magnetic encephalogram (MEG) and magnetic resonance image (MRI) of the subject's head. The inverse problem solutions, to be physiologically interpreted, must be plotted at the MRI. During the MEG recording, three fiducial markers are applied for the correct localization of the head. In this article, we use MEG and MRI for one control subject, obtained in the brain spontaneous activity study in the Center for Neuromagnetism of New York University School of Medicine [13, 14]. The MEG data were recorded with a 275-channel synthetic third order gradiometer (VSM MedTech LTD) in a magnetically shielded room, sampling rate was 1200 Hz. The NYU Institutional Review Board approved the study and an informed written consent was obtained from subject before the MEG recording. The subject was asked to relax but stay awake with eyes closed during 7-minute recording period.

Mathematical biology and bioinformatics. 2017. V. 12. № 1. doi: 10.17537/2017.12.176

Inverse problem solution

The inverse problem solution in MEG addresses the finding of the magnetic field sources from the known values of magnetic induction at some sensors on the head surface. To solve this problem, the following function depending on the magnetic field sources is minimized:

$$f = \sum_{i=1}^{N} \omega_i (B_i - B_i^0)^2 \to \min . \tag{1}$$

Here B_i^0 are the values of the magnetic induction measured by the sensors, B_i are the relevant values from forward field modeling, ω_i are the sensors' weights, and N is the number of sensors. In terms of spectral-based approach B_i^0 are the values of restored MEG channels at selected frequency at given moment of time. Provided with the initial guess, the dipoles location is determined b y standard mathematical methods designed for searching the local minimum of the function of several variables. Since in this case the information on the derivatives of the function being minimized is difficult to obtain, the zero-order methods were selected. Namely, the Nelder-Mead simplex method [25] is used for minimization. As forward field model, we use equivalent current dipole model in conducting sphere [26]. The results of this procedure are dipole coordinates and direction. These results are shown superimposed at subject's MRI. In this article, the inverse problem was solved from the field maps, extracted by methods of the data analysis, presented at the MathBrain.

Fourier transform

Direct and inverse Fourier transforms are the basic methods for the analysis of MEG time-series. Multichannel magnetic recording can be presented as the set of discrete experimental vectors $\{B_l\}$, where l is the number of registration channel. The instantaneous field values $B_l(i)$ are registered at the time moments τ_i; $i = 0, \ldots, N-1$; $\tau_0 = 0$. Discrete-time Fourier transform of these vectors is defined as:

$$b_l(k) = \sum_{n=0}^{N-1} B_l(n) \cdot e^{-\frac{i2\pi kn}{N}}, \; k = 0, \ldots, N-1 . \tag{2}$$

Since our input vectors are real-valued, we use only first $N/2$ components of DFT. Phase and amplitude of corresponding sinusoidal component of Fourier series can be defined as:

$$\varphi_l(k) = atan2(Im(b_l(k)), Re(b_l(k))), \tag{3}$$

$$\rho_l(k) = \frac{\sqrt{Re(b_l(k))^2 + Im(b_l(k))^2}}{N}, \tag{4}$$

where $Re(b_l(k))$ is the real part of $b_l(k)$, and $Im(b_l(k))$ is the imaginary part. The frequencies of this transform are: $\nu_k = \frac{k}{T}$; $k = 0, \ldots, \left[\frac{N}{2}\right]$, where $T = \tau_{N-1}$ is the length of the recording.

The MathBrain service makes it possible to study the detailed frequency structure of the brain, restoring multichannel signal at every frequency and analyzing the field maps obtained. The multichannel signal is restored at the particular frequency in all channels:

$$B_{lk}(n) = \rho_l(k)\,sin\left(\frac{2\pi kn}{N} + \varphi_l(k)\right). \tag{5}$$

The summarized instantaneous power produced by all channels is:

$$p_k(n) = \sum_{l=0}^{L} B_{lk}^2(n). \tag{6}$$

The field map for the inverse problem solution is selected at the moment of maximal power

$$B_{lk}(n_{max}),\, l = 1,\ldots,L. \tag{7}$$

Selection of the frequency, maximal power field map and inverse problem solution are illustrated in Fig. 2.

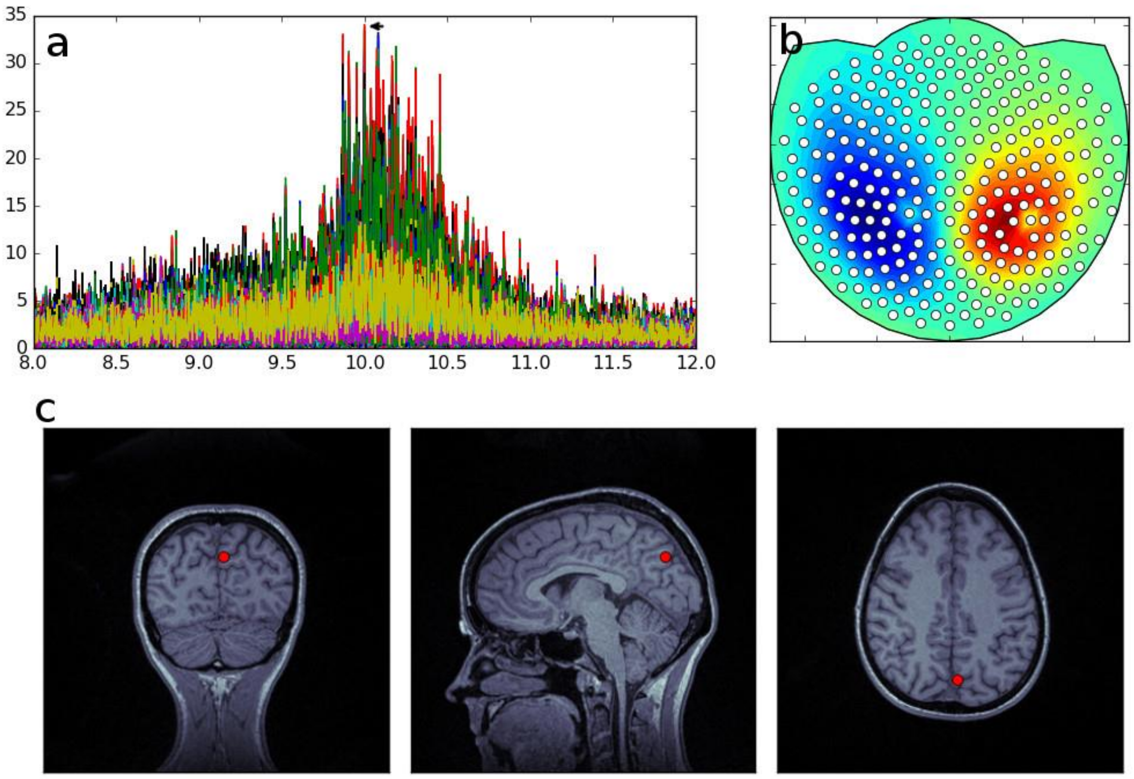

Fig. 2. a – Multichannel spectrum in the alpha-rhythm frequency band. Frequency to be analyzed is shown by arrow. **b** – Magnetic field map for the moment of maximal power. **c** – Inverse problem solution (red circle) shown at the MRI tomographic sections.

To calculate the direct and inverse Fourier transforms we are using Python wrapper for FFTW3 library [23, 24]. User can specify output frequency band, in this case all coefficients, corresponding to frequencies lower than user-selected lower bound are filled with zeroes, frequencies higher than upper-bound are truncated. Inverse Fourier transform is used for restoration of time-series from spectrum. It can be computed on selected time-frame or on the whole length of the original time-series. Also at this stage one can select frequency band for restoration. Restored time-series can be used in further PCA and ICA analysis.

Karhunen-Loeve transform

Discrete Karhunen-Loeve transform [27, 28] or principal component analysis is multivariate signal processing technique widely used in neuroimaging to decompose a multivariate dataset in a set of successive orthogonal components that explain a maximum amount of the variance. Let $\mathbf{B} = (b_1\ b_2\ \ldots\ b_N)^T$, were N is number of channels, b_n is the experimental vector of the magnetic recordings in the n-th channel. Principal component transform of **B** is:

$$Y = \mathbf{\Phi}^T \mathbf{B}, \tag{8}$$

where $\mathbf{\Phi} = (\varphi_1\ \varphi_2\ \ldots\ \varphi_N)^T$ is the orthonormal matrix, consisting of φ_n – eigenvectors of the following covariance matrix:

$$\left[\Sigma_{ij}\right] = \mathbf{E}\left[B_i B_j\right],\ \forall i, j \in \{1, \ldots, N\}. \tag{9}$$

In MathBrain the Karhunen-Loeve transform is applied to multi-channel MEG time-series in “temporal” variant, i.e. variance is intended in the domain of temporal observations whereas the number of channels defines the original dimensionality of the multivariate data set. The principal components can be used as input field-maps for the inverse problem solution, as is illustrated in Fig. 3.

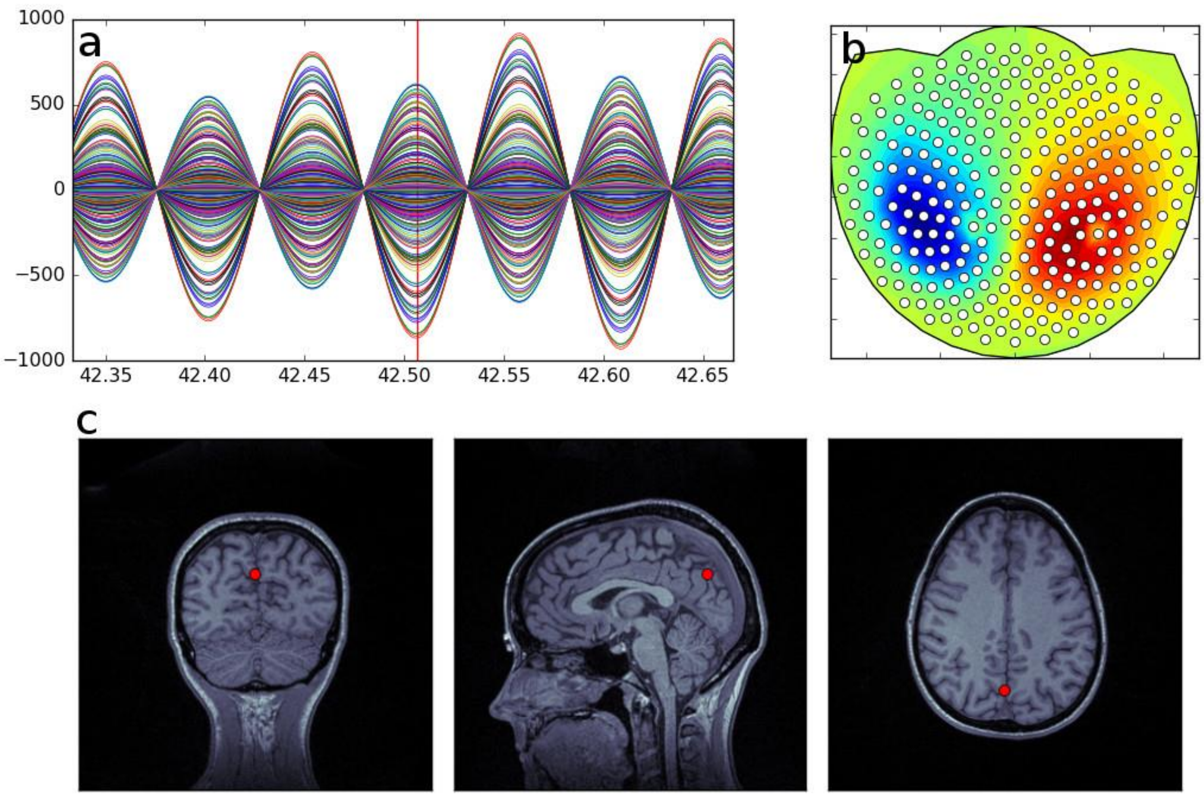

Fig. 3. a – Projection of the MEG at the first (most powerful) principal component. Time moment to be analyzed is shown by red cursor. **b** – Magnetic field map for the selected time moment. **c** – Inverse problem solution (red circle) shown at the MRI tomographic sections.

Independent Component Analysis

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. It is well known in neuroimaging and MEG data processing and is used to separate mixed signals. Let $\mathbf{B} = (b_1\ b_2 \ldots b_N)^T$, where N is number of channels, b_n is an experimental vector of the magnetic recordings in the n-th channel. Independent component decomposition for **B** is:

$$\mathbf{X} = \mathbf{W}\mathbf{B} \tag{10}$$

Where $\mathbf{X}$ is set of independent components and $\mathbf{W}$ is unmixing matrix. To calculate unmixing matrix we use FastICA algorithm [29], implemented in scikit-learn library [22]. From independent components and mixing matrix one can reconstruct multichannel time-series for each of independent components and solve the inverse MEG problem (see Fig. 4).

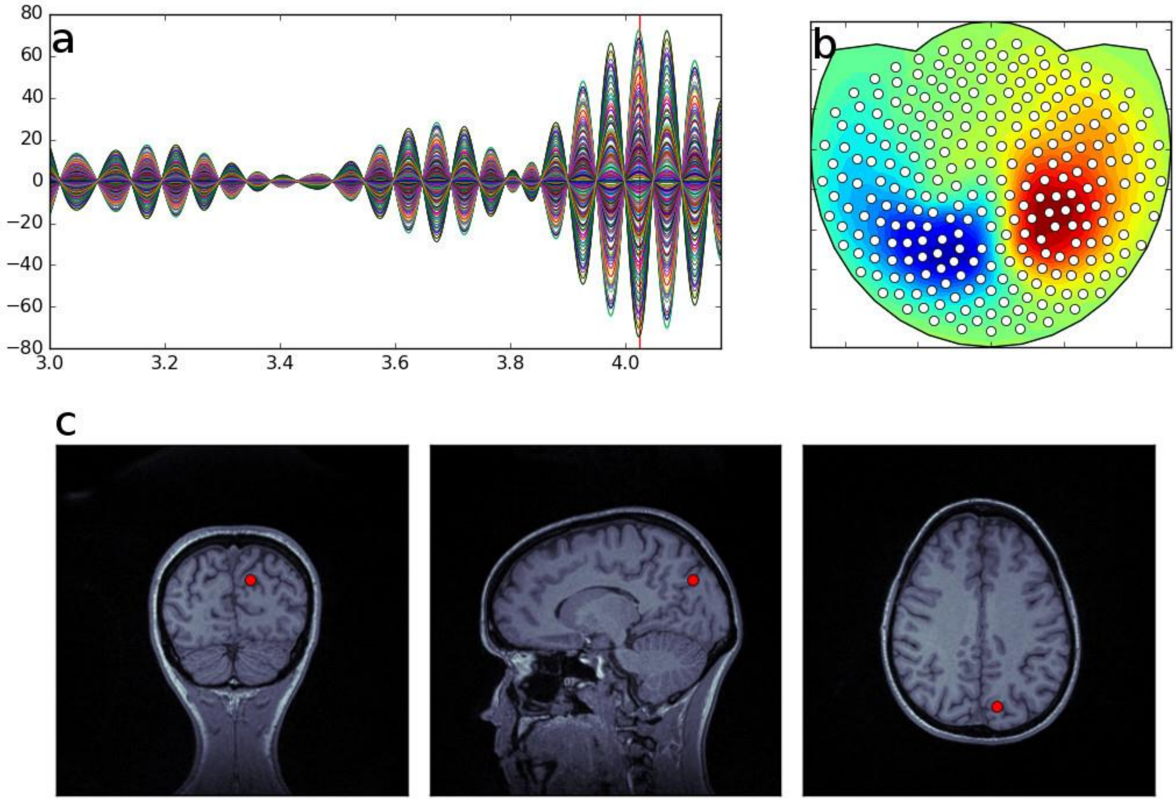

Fig. 4. **a** – The most powerful independent component. Time moment to be analyzed is shown by red cursor. **b** – Magnetic field map for the selected time moment. **c** – Inverse problem solution (red circle) shown at the MRI tomographic sections.

Quantitative analysis

The quantitative analysis of encephalography data is performed by direct Fourier transform in user-defined moving time-window of MEG time-series. As the result of this analysis, one-dimensional power spectrogram (sum of powers in all channels in corresponding frequency bin) is produced, this is used to evaluate spectral changes of signal value during the time of measurement.

From the spectrogram one can select the time moment and frequency to be analyzed. Then the one-frequency MEG can be reconstructed, using (6), and the field map (7) can be selected for inverse problem solution (See Fig. 5).

Mathematical biology and bioinformatics. 2017. V. 12. № 1. doi: 10.17537/2017.12.176

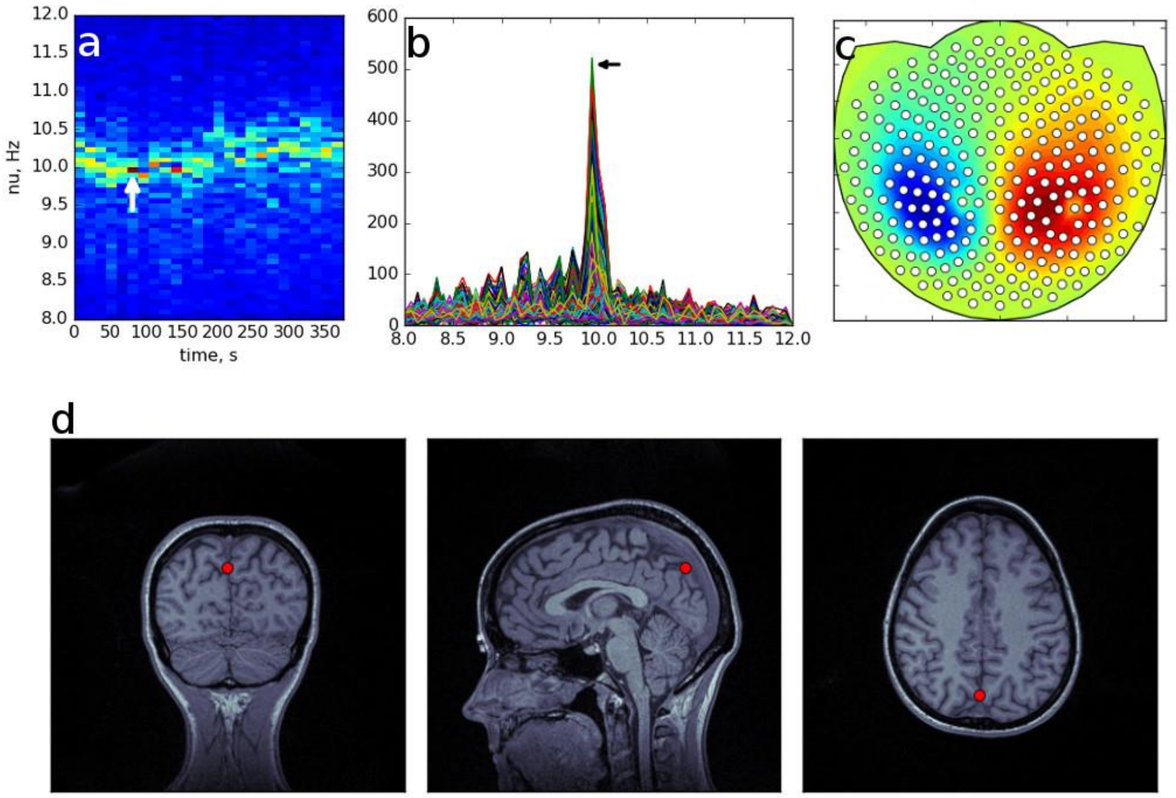

Fig. 5. a – The spectrogram, calculated in 15 second time-window. Time moment and frequency to be analyzed is shown by white arrow. **b** – multichannel spectrum for selected time moment. Frequency to be reconstructed is shown by black arrow. **c** – Magnetic field map for the reconstructed MEG. **d** – Inverse problem solution (red circle) shown at the MRI tomographic sections.

CONCLUSIONS

The cloud service MathBrain was developed, providing several MEG data processing methods. In this article, we have used these methods for the analysis of the real data set for the control subject, including MEG and MRI. The alpha rhythm frequency band (8–12 Hz) was selected, containing powerful broad peak for the «Eyes closed» condition of the subject (See Fig. 2,a). Using inverse Fourier transform, we have reconstructed MEG in the band 8–12 Hz, and further analyzed it using all methods of the MathBrain service.

The main problem of the multicomponent analysis methods is to find the criteria for component selection. Here we used the criterion of maximal power, in every method picking the strongest components for inverse problem solution. It can be concluded, that results of various methods are substantially similar, and generally correspond to present knowledge about the alpha rhythm localization. Variations between methods can be explained by the fact, that alpha rhythm is generated by many sources, covering rather broad space (5–8 cm), see [14].

Further development of the MathBrain service will include methods of integrated analysis, such as Functional Tomography [12–14], Partial Spectroscopy [15] and connectivity analysis.

This work was partially supported by the Russian Foundation for Basic Research (grants 16-07-00937, 16-07-01000, 17-07-00677, 17-07-00686), by the Program I.33P for Fundamental Research of the Russian Academy of Sciences, and by the CRDF Global (USA) (grants CRDF RB1-2027 and RUB-7095-MO-13).

REFERENCES

1. Dietsch G. Fourier-analyse von elektroenzephalogrammen des menschen. *Pflügers Arch. Ges. Physiol*. 1932. V. 230. P. 106–112. doi: 10.1007/BF01751972
2. Jansen B.J., Bourne J.R., Ward J.W. Spectral decomposition of EEG intervals using Walsh and Fourier transforms. *IEEE Trans. Biomed. Eng*. 1981. V. 28. P. 836–838. doi: 10.1109/TBME.1981.324686
3. Halliday D.M., Rosenberg J.R., Amjad A.M., Breeze P., Conway B.A., Farmer S.F. A frame work for the analysis of mixed time series point process data–theory and application to study of physiological tremor, single motor unit discharges and electromyogram. *Prog. Biophys. Mol. Biol*. 1995. V. 64. P. 237–238. doi: 10.1016/S0079-6107(96)00009-0
4. Cooley J.W., Tukey J.W. An algorithm for the machine calculation of complex fourier series. *Math. Comput*. 1965. V. 19. P. 297–301. doi: 10.1090/S0025-5718-1965-0178586-1
5. Miyashita T., Ogawa K., Itoh H., Arai H., Ashidagawa M., Uchiyama M., Koide Y., Andoh T., Yamada Y. Spectral analyses of electroencephalography and heart rate variability during sleep in normal subjects. *Auton. Neurosci*. 2003. V. 103. P. 114–120. doi: 10.1016/S1566-0702(02)00259-X.
6. Weiss S., Rappelsberger P. Long-range EEG synchronization during word encoding correlates with successful memory performance. *Cogn. Brain Res*. 2000. V. 9. P. 299–312. doi: 10.1016/S0926-6410(00)00011-2
7. Jarvis M.R., Mitra P.P. Sampling properties of the spectrum and coherency of sequences of acton potentials. *Neural Comput*. 2001. V. 13. P. 717–749. doi: 10.1162/089976601300014312
8. Delorme A., Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods*. 2004. V. 134. P. 9–21. doi: 10.1016/j.jneumeth.2003.10.009
9. Garcia-Rill E., Moran K., Garcia J., Findley W.M., Walton K., Strotman B., Llinas R.R. Magnetic sources of the M50 response are localized to frontal cortex. *Clin. Neurophysiol*. 2008. V. 119. P. 388–398. doi: 10.1016/j.clinph.2007.10.027
10. Muthuraman M., Galka A., Deuschl G., Heute U., Raethjen J. Dynamical correlation of non-stationary signals in time domain – a comparative study. *Biomed. Signal Process. Control*. 2010. V. 5. P. 205–213. doi: 10.1016/j.bspc.2010.02.006
11. Llinás R.R. *I of the Vortex. From Neurons to Self*. Cambrige, MA: MIT Press, 2001.
12. Llinás R.R., Ustinin M.N. *Precise Frequency-Pattern Analysis to Decompose Complex Systems into Functionally Invariant Entities*: U.S. Patent. US Patent App. Publ. 20160012011 A1. 01/14/2016.
13. Llinas R.R., Ustinin M.N. Frequency-pattern functional tomography of magnetoencephalography data allows new approach to the study of human brain organization. *Frontiers in Neural Circuits*. 2014. V. 29. No. 8. P. 43. doi: 10.3389/fncir.2014.00043
14. Llinas R.R., Ustinin M.N., Rykunov S.D., Boyko A.I., Sychev V.V., Walton K.D., Rabello G.M. and Garcia J. Reconstruction of human brain spontaneous activity based on frequency-pattern analysis of magnetoencephalography data. *Frontiers in Neuroscience*. 2015. V. 9. P. 373. doi: 10.3389/fnins.2015.00373
15. Rykunov S.D., Ustinin M.N., Polyanin A.G., Sychev V.V., Llinás R.R. Software for the Partial Spectroscopy of Human Brain. *Mathematical Biology and Bioinformatics*. 2016. V. 11. № 1. P. 127–140. doi: 10.17537/2016.11.127 (in Russ.)
16. Ustinin M.N., Sychev V.V., Walton K.D., Llinas R.R. New Methodology for the Analysis and Representation of Human Brain Function: MEGMRIAn. *Mathematical Biology and Bioinformatics*. 2014. V. 9. № 2. P. 464–481. doi: 10.17537/2014.9.464.

17. Oostenveld R., Fries P., Maris E., Schoffelen J.M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*. 2011. V. 2011. doi: 10.1155/2011/156869
18. Delorme A., Mullen T., Kothe C., Akalin Acar Z., Bigdely-Shamlo N., Vankov A., Makeig S. EEGLAB, SIFT, NFT, BCILAB, and ERICA: New Tools for Advanced EEG Processing. *Computational Intelligence and Neuroscience*. 2011. V. 2011. Article No. 130714. doi: 10.1155/2011/130714
19. Tadel F., Baillet S., Mosher J.C., Pantazis D., Leahy R.M. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*. 2011. V. 2011. Article No. 879716. doi: 10.1155/2011/879716
20. Oplachko E.S., Ustinin D.M., Ustinin M.N. Cloud Computing Technologies and their Application in Problems of Computational Biology. *Mathematical Biology and Bioinformatics*. 2013. V. 8. № 2. P. 449–466. doi: 10.17537/2013.8.449 (in Russ.)
21. Oliphant T.E. Python for Scientific Computing, *Computing in Science & Engineering*. 2007. V. 9. P.10–20. doi: 10.1109/MCSE.2007.58
22. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011. V. 12. P. 2825–2830.
23. *Python wrapper for FFTW3 library*. URL: https://github.com/pyFFTW/pyFFTW (accessed 02.05.2017).
24. Frigo M., Johnson S.G. The Design and Implementation of FFTW3. *Proceedings of the IEEE*. 2005. V. 93. P. 216–231. doi: 10.1109/JPROC.2004.840301
25. Nelder J.A., Mead R. A Simplex Method for Function Minimization. *The Computer Journal*. 1965. V. 7. P. 308–313. doi: 10.1093/comjnl/7.4.308
26. Sarvas J. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine and Biology*. 1987. V. 32. P. 11–22. doi: 10.1088/0031-9155/32/1/004
27. Karhunen K. Uber lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys*. 1947. № 37. P. 1–79.
28. Loeve M. Probability theory II. Springer-Verlag, 1978. (Graduate Texts in Mathematics, V. 46).
29. Hyvärinen A., Oja E. Independent component analysis: algorithms and applications. *Neural Networks*. 2000. V. 13. № 4–5. P. 411–430. doi: 10.1016/S0893-6080(00)00026-5

Received 04 April 2017.
Published 02 May 2017.