VirtualBrainCloud
Personalized Recommendations for Neurodegenerative Disease

www.VirtualBrainCloud-2020.e

Public deliverable report

D3.4: Initial MEG and SEEG brain dynamic measures at the disposal of other WPs.

Date	June 2020
Authors	Ehtasham Javed, Sheng H. Wang, Felix Siebenhühner, J. Matias Palva, Satu Palva (UNIVERSITY OF HELSINKI, UH – LEAD), Fernando Maestú, Isabel Suárez, Gianluca Susi (UNIVERSIDAD COMPLUTENSE DE MADRID, UCM) Gabriele Arnulfo (UNIVERSITA DEGLI STUDI DI GENOVA, UNIGE) Nikolaus Forgó , Michael Cepic, Mariana Rissetto (UNIVERSITY OF VIENNA, UNIVIE), Petra Ritter (CHARITE) © VirtualBrainCloud consortium
Dissemination level:	public
Website	www.VirtualBrainCloud-2020.eu

Table of content

1.   Abstract

This report is an outcome of the Virtual Brain Cloud addressing*: Data Processing, Standardization and Data-FAIRification.* It describes the initial pilot dataset containing electrophysiological spatio-temporal data of brain dynamics from healthy controls and patients aiming to support The Virtual Brain (TVB) parameter validation.

Purpose of data sharing in TVB-Cloud: Provide MEG/SEEG-based electrophysiological measures of spatiotemporal brain dynamics for multimodal validation of personalized TVB parameters.

Data Description: All data are pseudonymized. The interim dataset consists of brain data from 20 Magnetoencephalograms (MEG), 5 from each diagnosis group: Healthy Controls (HC), adults with Subjective Cognitive Decline (SCD) and Mild Cognitive Impairment (MCI) and patients with Alzheimer's Disease (AD) and one Stereoelectroencephalogram (SEEG) from one patient with epilepsy. Data are stored following the Brain Imaging Data Structure (BIDS).

-    **Data Derivatives:** MEG derivatives resulting from the pre-processing and source reconstruction pipelines are provided, including parcel time-series. We also provide TVB-compliant data for one of the MEG subjects and the SEEG subject.

-    **Source data:** The source data for MEG subjects have been reconstructed using the Linearly Constrained Minimum variance (LCMV), and considered for a wide-bandwidth of [2-45 Hz], along with bands: [4-8] (theta), [8-12] (alpha), [12-30] (beta), [12-20] (lower beta), [20-30] (upper beta), [30-45] (gamma). After this step,the data was aggregated in Regions of Interest, (ROIs) using different parcellation atlases which include Shaefer 400 parcels (Shaefer et al., 2018), Desikan-killiany (Desikan et al., 2006).

Research Results: Besides datasets, we also provide the results of brain dynamic measures for both MEG and SEEG dataset, which include: phase autocorrelation, long range temporal correlations, bistability index, Phase Locking Value (PLV), Corrected Imaginary Phase Locking Value (ciPLV), weighted Phase Lag Index (wPLI), Amplitude Envelope Correlation (AEC), phase synchronization of amplitudes, cross-frequency coupling, phase lags and phase difference distributions.

The results for analyses of SEEG are for a single subject. Section 5 provides an overview of each file in SEEG-BIDS.

Data sharing: To each subject we assigned a numerical code according to BIDS-standards. Patient to subject-id association is kept in a password-protected file which is safely stored in compliance with National and European Laws. The deliverable data are currently placed at local repositries of UCM (MEG) and UNIGE (SEEG), and are available to members of TVB-Cloud consortium after establishing data processing agreement. Note that for interim datasets, as per sharing agreement, we are prevented from sharing raw data, but derivatives are allowed to be shared. On the other hand, for SEEG data, we shall only share statistical derivatives (brain dynamic measures) according to current data sharing agreement.

These derivatives, in our view, represent an invaluable resource for optimal optimization of model parameters.

1.1. Subtitle

1.1.1. Sub-subtitle

2.   Partners involved

1 UNIVERSITY OF HELSINKI (UH) - LEAD
2 UNIVERSIDAD COMPLUTENSE DE MADRID (UCM)
3 UNIVERSITA DEGLI STUDI DI GENOVA (UNIGE)
4 UNIVERSITY OF VIENNA (UNIVIE)
5 CHARITÈ UNIVERSITY MEDICINE (CHARITE)

3.   Group Meetings

- VirtualBrainCloud Electrophysiology Group meeting, May 8th, 2019, Helsinki
  - o   Participant: UH, UCM, UNIGE
- VirtualBrainCloud Electrophysiology Group meeting, September 5th, 2019, Hilton Diagonal Mar Barcelona
  - o   Participant: UH, UCM, UNIGE
- Video Conference I to dicuss case specific legal aspects of data sharing, January 17th, 2020
  - o   Participants: UNIVIE, CHARITE, UH, UCM, UNIGE
- Video Conference II to discuss case specific aspects of data sharing | April 27th, 2020.
  - o   Participants: UNIVIE, UH, UCM, UNIGE
- VirtualBrainCloud Electrophysiology group + CHARITE virtual meeting | May 20th, 2020.
  - o   Participants: CHARITE, UH, UCM, UNIGE

4.   Overview of MEG-BIDS Repository

4.1.   Current limitations of BIDS for Magnetoencephalography and BIDS Extensions Proposals

BIDS (Brain Imaging Data Structure) for MEG was introduced in 2018 with the publication of *MEG-BIDS, the brain imaging data structure extended to magnetoencephalography*, by Niso and colleagues[1], , and the associated specification for the standardization of MEG data. This guideline covered the organization of raw and at most minimally processed data (e.g. after tSSS or in-machine derived data). After MEG-BIDS was introduced, open teams of researchers have been gathered under different BIDS Extension Proposals (BEPs) to expand the specification in a community-driven effort[2]. The BEPs are works in draft stage, subjected to ongoing modifications by the teams and the community working on them. In order to provide a complete representation of the outcomes generated in the processing of MEG data up to source-space and functional connectivity metrics, we have introduced a customized organization of derived data, following where possible the proposals from the BEPs and the original BIDS specification. In this regard, we adhered particularly to: 1) *BIDS Extension Proposal 03 (BEP003): Common Derivatives* and 2) *BIDS Extension Proposal 21 (BEP021): Common Electrophysiological Derivatives* in their current form (November 2019). The BEPs leave space for customization of labels, variables and metadata, given the variety of methods and pipelines that are used within the MEG research community. However, even though it is hoped that the procedure we adapted in organizing the data for this project can be of use to the developers of the MEG-BIDS community, it is to be expected that for as long as MEG-BIDS is a work in progress, our BIDS dataset will remain a work in progress that should be updated whenever a consensus is reached within the MEG-BIDS community.

Another issue arises from the large size of the MEG-derived data (approximately 5.5 GB per subject in our case), and the lack of consensus regarding data formats. Following BIDS, MEG-BIDS favours text formats such as Tab Separated Value (.tsv) for tabular data and JavaScript Object Notation (.json) format for key/value files and acknowledges manufacturer-specific extensions for raw data (.fif, .ds, .sqd, .con, .raw and so on). The analysis of MEG data commonly requires handling heavy matrices of data that are often n-dimensional, rendering cumbersome and suboptimal their conversion to tabular text. BEP021 acknowledges this issue and lists the pros and cons of different data formats to be used whenever this problem is met (.mat up to Version 7.3, .npy, .txt and .h5). For convenience, we have maintained the .mat format in those particular cases as derived from our pipeline, consistently attaching extra documentation for clarity. Also, the storage of source-space time-series required MAT-files of Version 7.3 and thus producing computationally-demanding large files (see **Table 1**). In this case, we stored the interim source-space data series in a binary .dat format providing the code for .mat conversion. Once consensus is reached, we will adapt our interim database to the convenient data formats agreed upon.These choices attempted to provide a plausible and reasonable representation of the MEG metadata (see **Section 4.4**) that is required for most common data processing operations and as generated by a classical processing pipeline. However, our decisions are strongly based on our

---

[1] Niso, G., Gorgolewski, K., Bock, E. et al. MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. Sci Data 5, 180110 (2018). https://doi.org/10.1038/sdata.2018.110
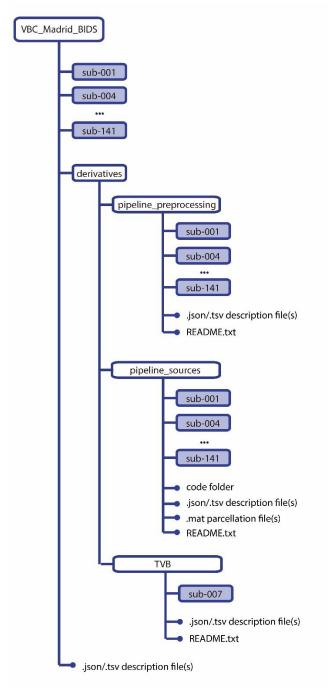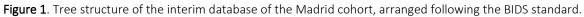[2] https://bids-specification.readthedocs.io/en/latest/06-extensions.html

language of choice and on the outputs of specific processing pipeline and will in all cases be adapted for compatibility with the TVB-Cloud applications whenever necessary.

4.2. An outlook

The interim MEG-BIDS database is structured following the BIDS standard, and as such, it is composed of a **main folder** as defined in the specification, and a **derivatives folder** comprising the derived data generated in the different processing pipelines (/**pipeline_preprocessing** and /**pipeline_sources**). Additionally, it also comprises a set of **TVB-compliant** structural and functional connectivity data (/**TVB**). Within each processing pipelines folder, the BIDS structure of the main folder is replicated yielding one subfolder per subject. A README file found at the root of each processing pipelines folder provides detailed information to understand the derived data contained therein (see **Figure 1**).

Figure 1. Tree structure of the interim database of the Madrid cohort, arranged following the BIDS standard.

4.3. Size of the data

For each folder (main folder, derivatives/pipeline_preprocessing, derivatives/pipeline_sources and derivatives/TVB), folder size is estimated for an example subject 'sub-00n':

Table 1: Data sizes per subject and subfolder for a complete subject of the interim MEG database.

VBC_Madrid_BIDS/sub-<participant_label>/	35 MB approx.
VBC_Madrid_BIDS/derivatives/pipeline_preprocessing/sub-<participant_label>/	1.5 GB approx.
VBC_Madrid_BIDS/derivatives/pipeline_sources/sub-<participant_label>/	4 GB approx.
VBC_Madrid_BIDS/derivatives/TVB/sub-<participant_label>/	50 MB approx.
TOTAL data size for each subject	**5.5 GB approx.**

For this deliverable, we provide the complete dataset for one of the subjects (sub-007), in order to show inputs and outputs of each processing block adopted during the data processing. Regarding the other subjects, the missing data is specified with an asterisk in tables 3,4,5.

4.4. Summary of the files

4.4.1. Main Folder

The **main folder** contains the information relative to relevant MEG data acquisition parameters and manufacturer information as envisaged in the MEG-BIDS proposal (subfolder **/meg**). Additionally, this folder hosts **MRI data** for each subject in the database (defaced T1-weighted MRI and relevant acquisition and manufacturer parameters) (subfolder **/anat**). A detailed description of the metadata can be found in the original MEG-BIDS specification. The files found under the main folder are briefly summarized in **Table 2**.

Table 2: Data provided for an example subject in the **main folder** of the interim MEG database.

VBC_Madrid_BIDS/sub-<participant_label>/ses-<session_label>/**meg**	
Filename	**Description**
sub-<participant_label>_ses-<session_label>_coordsystem.json	This file contains the information relative to the MEG coordinate system.
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_channels.tsv	This file contains the information relative to the MEG sensors (type, units, sampling frequency).
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_meg.json	This file contains the information relative to the relevant acquisition parameters (sampling frequency, anti-aliasing filtering, institution, instructions given to the participants, etc.).
VBC_Madrid_BIDS/sub-<participant_label>/ses-<session_label>/**anat**	

sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_T1w.nii	This file contains the defaced T1-weighted MRI in .nii format.
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_T1w.json	This file contains the information relative to the relevant parameters of the MRI acquisition protocol.

4.4.2. Derivatives subfolder

The **derivatives subfolder** hosts the information relative to the derived data as generated in the different stages of the processing of MEG data (see Sections **6.1.3.** and **6.1.4.**). Specifically, the subfolder **/pipeline_preprocessing** (see **Table 3**) stores the data derived from MEG preprocessing, and the subfolder **/pipeline_sources** (see **Table 4**) follows with the information generated in the source-reconstruction pipeline. Both pipelines were run on Matlab using the Fieldtrip Toolbox[3]. A detailed description of the content of each derivatives subfolder is provided in a README file found at the root of each subfolder:

- VBC_Madrid_BIDS/derivatives/pipeline_preprocessing/README.txt.
- VBC_Madrid_BIDS/derivatives/pipeline_sources/README.txt.
- VBC_Madrid_BIDS/derivatives/TVB/README.txt.

Table 3: Data provided in the folder 'VBC_Madrid_BIDS/derivatives/pipeline_preprocessing' of the interim MEG database. We indicate the data currently available for sub-007 only with an asterisk preceding the filename.

VBC_Madrid_BIDS/derivatives/**pipeline_preprocessing**/ sub-<participant_label>/	
Filename	**Description**
*sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-**SOBI_mixing**.tsv*	This table contains the mixing matrix for the Independent Component Analysis performed using Second Order Blind Identification (SOBI) algorithm. Rows represent the MEG channels; columns represent the Independent Components.
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_mixing.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_mixing.tsv.*
*sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-**SOBI_unmixing**.tsv*	This file contains the pseudo-inverse of the mixing matrix for the Independent Component Analysis performed using Second Order Blind Identification (SOBI) algorithm. Rows represent the Independent Components; columns represent the MEG channels.
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_unmixing.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_unmixing.tsv.*

[3] Oostenveld, R., Fries, P., Maris, E., Schoffelen, JM. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience 2011, 156869 (2011). doi:10.1155/2011/156869

** sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-**epochs_meg**.mat*	This file contains the matrix (epochs x channels x timepoints) representation of the segmented MEG data in Matlab's .mat format. Epochs of 4 seconds with 2 seconds of padding (real data) on both edges.
* sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-epochs_meg.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-epochs_meg.mat* in addition to the required and recommended fields in the _meg.json file of the BIDS core specification.
*sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_**epochs**.tsv*	This file contains the temporal definition of the epochs. The following CustomMetadata fields are provided: - 'duration'. Duration of the segmented epochs. In seconds. - 'beg_time'. Beginning time. Time with respect to the start of the recording when the current epoch starts regardless of the padding. In seconds. - 'end_time'. End time. Time with respect to the start of the recording when the current epoch ends regardless of the padding. In seconds. - 'beg_padd'. Beginning time with padding. Time with respect to the start of the recording when the current epoch starts with added padding. In seconds. - 'end_padd'. End time with padding. Time with respect to the start of the recording when the current epoch ends with added padding. In seconds. - 'padd_length'. Padding length. Length of the padding that was added. In seconds. - 'trial_type'. Description of the type of trial.
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_epochs.json	This file provides the information required for *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_epochs.tsv*
*sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-**recording_annotations**.tsv*	This table contains the information relative to the annotations of the time interval corresponding to the eyes-closed resting state condition. Only this time interval was used for further pre-analyses. - *'onset'*. Onset in seconds from the beginning of the recording. 'n/a' if it affects the entire recording. If 'n/a', duration must also be 'n/a'. - *'duration'*. Duration in seconds. 0 for instantaneous events. 'n/a' if it affects the entire recording. If 'n/a', onset must also be 'n/a'. - *'label'*. Label for the annotation. It can be a list of annotations separated by comma: e.g. 'rest_eyesclosed' for eyes-closed resting-state. - *'channels'*. Label specific to a channel. If 'n/a', applies to all the channels.
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-recording_annotations.json	This file contains the information relative to the description of the file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-recording_annotations.tsv*.

sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-**artefacts_annotations**.tsv	This table contains the information relative to the annotations of the resulting definition of the remaining artefacts after completion of the pre-processing pipeline (including ICA component removal). It includes the information (onset, duration and type) of the artefacted segments (eye_blink, muscle, flux_jumps). *'onset'*. Onset in seconds from the beginning of the recording. 'n/a' if it affects the entire recording. If 'n/a', duration must also be 'n/a'. *'duration'*. Duration in seconds. 0 for instantaneous events. 'n/a' if it affects the entire recording. If 'n/a', onset must also be 'n/a'. *'label'*. Label for the annotation. It can be a list of annotations separated by a comma: - 'eye_blink'. Ocular artefact caused by blinking or eye movement. - 'flux_jump'. SQUID jump artefact. - 'muscle'. Muscular artefact caused by muscular contraction or swallowing. 'channels'. Label specific to a channel. If 'n/a', applies to all the channels.
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-artefacts_annotations.json	This file contains the information relative to the description of the file sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-artefacts_annotations.tsv.

sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-**SOBI_annotations**.tsv	This table contains the information relative to the annotations of the resulting definition of undesirable components after Independent Component Analysis (ICA) in both time and frequency domains.
	It includes the information (type and name of ICA components) of the undesirable components.
	'onset'. Onset in seconds from the beginning of the recording. 'n/a' if it affects the entire recording. If 'n/a', duration must also be 'n/a'.
	'duration'. Duration in seconds. 0 for instantaneous events. 'n/a' if it affects the entire recording. If 'n/a', onset must also be 'n/a'.
	'label'. Label for the annotation. Can be a list of annotations separated by comma:
	  - 'eye_blink'. ICA component caused by blinking or eye movement.
	  - 'heart_beat'. ICA component of cardiac activity.
	  - 'noise'. ICA component of undefined noise.
	'*channels*'. Label specific to a channel. If 'n/a', applies to all the channels. Else, it must be a list of channel names or labels for the time series (e.g., name of ICA component).

sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_annotations.json	This file contains the information relative to the description of the file sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_annotations.tsv.

Table 4: Data provided in the folder 'VBC_Madrid_BIDS/derivatives/pipeline_sources' of the interim MEG database. We indicate the data currently available for sub-007 only with an asterisk preceding the filename.

VBC_Madrid_BIDS/derivatives/**pipeline_sources**/sub-<participant_label>/	
Filename	**Description**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_**headmodel**.zip	This compressed folder contains the data relative to the triangulated surface mesh used to provide geometrical description of the volume conduction model of the head . Coordinates are given in Electa-Neuromag space. The following files are provided in TVB format: - 'triangles.txt'. Matrix of (triangles x 3) indicating the combinations of indexes from the vertices.txt array forming each triangle. - 'vertices.txt'. Matrix of (NumberOfVertices x 3) indicating the coordinates (x, y, z) of each vertex in a space separated values file. - 'normals.txt'. Matrix of (NumberOfVertices x 3) indicating the normal vector (x, y, z) to each vertex in a space separated values file.
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_headmodel.json	This file contains the details of the parameter combinations used to generate the derived file sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_headmodel.zip. The following CustomMetadata fields are provided: - 'HeadmodelType'. Volume conduction model, e.g. 'singleshell', 'localspheres'. - 'Tissue'. Tissue that is being modeled, e.g. 'brain'. - 'Unit'. Measure unit (International System units). - 'NumberOfVertices'. Number of vertices used to build the geometrical headmodel. - 'FTCode'. Fieldtrip script used to compute the headmodel - 'FTMethod'. Fieldtrip method used in the call to FTCode.
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_**sourcemodel**.tsv	This table contains the sourcemodel used for source reconstruction of the MEG data. Provides a description of the Elekta-Neuromag coordinates of the equivalent current dipoles considered for source reconstruction, expressed as a matrix of dimensions (NumberOfSources x 4): - 'pos1'. Coordinate 1 of the position of the source. In meters. - 'pos2'. Coordinate 2 of the position of the source. In meters. - 'pos3'. Coordinate 3 of the position of the source. In meters. - 'inside'. Logical. 1 = Source inside of the headmodel. 0 = Source outside of the headmodel. Logical.
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_sourcemodel.json	This file contains the details of the parameter combinations used to generate the derived file sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_sourcemodel.tsv The following CustomMetadata fields are provided: - 'NumberOfSources'. Number of sources in the sourcemodel.

	- 'GridSourceDimension'. Number of sources along each coordinate axis. - 'SourceOrientation'. Vector indicating the orientation of the equivalent current dipoles.
*sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_**leadfield**.mat*	This file contains the forward solution in the volume conduction model, expressed as the leadfield matrix (sources x channels x 3) in Matlab's .mat format. 'n/a' is associated with sources that are not inside the headmodel (see the field "inside" in *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_sourcemodel.tsv*).
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_leadfield.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_leadfield.mat.*
*sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-<band_label>_**lcmvfilter**.tsv*	This table contains the beamformer filters to estimate the activity at any given location in the brain, computed using Linearly Constrained Minimum Variance (LCMV) method that minimizes the source variance and is subjected to a unit-gain constraint. Data is expressed as a matrix of dimensions (sources x channels). 'n/a' is associated with sources that are not inside the headmodel (see the field 'inside' in *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_sourcemodel.tsv*).
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-<band-label>_lcmvfilter.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-<band_label>_lcmvfilter.tsv.* The following CustomMetadata fields are provided: - 'Method'. Method used for the computation of the inverse solution. - 'FrequencyBand'. Name of the frequency band, e.g. 'Alpha', 'Beta', 'Broadband'. - 'FrequencyLimits'. Frequency limits of the frequency band, e.g. [8 12], [12 30], [2 45]. In Hertz.
*sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-<band_label>_**lcmvpow**.tsv*	This table contains the estimated power of the neuronal sources for FrequencyBand. 'n/a' is associated with sources that are not inside the headmodel (see the field "inside" in *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_sourcemodel.tsv*).
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-<band_label>_lcmvpow.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-<band_label>_lcmvpow.tsv.*
* *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-**sources_meg**.dat*	This file contains the matrix (epochs x sources x timepoints) representation of the source-space reconstructed time-series. Epochs of 4 seconds with added padding of PaddingLength on both edges. Data is stored in binary format and can be recovered to .mat formatting using the code provided in *sub-*

	<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-sources_meg.json CustomMetadata.CodeURL.
* sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-sources_meg.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-sources_meg.dat* in addition to the required and recommended fields in the _meg.json file of the BIDS core specification. The following CustomMetadata fields are provided: - 'MatrixDimensions'. Dimensions of the matrix of source-space reconstructed time-series: [epoch's sources time-points]. Required to recover the data with CodeURL. - 'FrequencyBand'. Name of the frequency band in which sourcedata is reconstructed, e.g. 'Alpha', 'Beta', 'Broadband'. 'FrequencyLimits'. Frequency limits of the frequency band, e.g. [8 12], [12 30], [2 45]. In Hertz. - 'Padding'. 'Yes' / 'No'. - 'PaddingLength'. Length of the padding that was added to the data (in seconds). - 'CodeURL'. Matlab code for the retrieval of the sourcedata matrix from its binary input *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-sources_meg.dat.*
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_atlas-<atlas_label>_meg**.dat*	This file contains the matrix (epochs x rois x timepoints) representation of the source-space parcel time-series. Epochs of 4 seconds with added padding of PaddingLength on both edges. Data is stored in binary format and can be recovered to .mat formatting using the code provided in *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_atlas-<atlas_label>_meg.json* CustomMetadata.CodeURL.
*sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_atlas-<atlas_label>_meg.json	This file contains the details of the parameter combinations used to generate the derived file sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_atlas-<atlas_label>_meg.dat in addition to the required and recommended fields in the _meg.json file of the BIDS core specification. The following CustomMetadata fields are provided: - 'MatrixDimensions'. Dimensions of the matrix of source-space parcel time-series: [epochs rois timepoints]. Required to recover the data with CodeURL. - 'Atlas'. Parcellation atlas used to compute the parcel time-series. - 'FrequencyBand'. Name of the frequency band in which sourcedata is reconstructed, e.g. 'Alpha', 'Beta', 'Broadband'. - 'FrequencyLimits'. Frequency limits of the frequency band, e.g. [8 12], [12 30], [2 45]. In Hertz. - 'Padding'. 'Yes' / 'No'. - 'PaddingLength'. Length of the padding that was added to the data. In seconds. - 'CodeURL'. Matlab code for the retrieval of the sourcedata matrix from its binary input *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_atlas-<atlas_label>_meg.dat.*

Additionally, 2 extra files provide relevant information about the **parcellation** and can be found at the root of the **pipeline_sources** subfolder:

- desikan.mat (for the Desikan-Killiany parcellation).
- schaefer400.mat (for the Schaefer 400 parcellation).

These parcellation files allow the association between the sources and the region indices for the respective anatomical parcellations.

4.4.3. TVB subfolder

Finally, the dataset contains a **TVB subfolder** including TVB-compliant data such as structural (SC) and functional (FC) connectivity data (see **Table 5**).

Table 5: Data provided in the folder 'VBC_Madrid_BIDS/derivatives/TVB/' of the interim MEG database. The asterisk before the filename indicates that the data is currently available for sub-007 only.

VBC_Madrid_BIDS/derivatives/**TVB**/sub-<participant_label>	
Filename	**Description**
** sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_**FC**.mat*	This file contains the Functional Connectivity (FC) data for the current subject, expressed as a matrix of dimensions (rois x rois) in Matlab's .mat format.
* sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_FC.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_FC.mat*. The following CustomMetadata fields are provided: - 'FrequencyBand'. Name of the frequency band (e.g., 'Alpha', 'Beta', 'Broadband'). - 'FrequencyLimits'. Frequency limits of the frequency band (e.g., [8 12], [12 30], [2 45]). In Hertz. - 'Metric'. FC metric used (e.g,. 'Phase-Locking Value', 'Phase-Lag Index', etc.). - 'Atlas'. Parcellation atlas used to parcel the FC.
** sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_**SC**.zip*	This compressed file contains the Structural Connectivity (SC) data for the current subject. The following CustomFiles files are provided in TVB format: - 'centres.txt'. Matrix (rois x 4) of SC centres. Each row corresponds to the center coordinates of a region. The 1st column is the region label. 2nd to 4th columns are the (x, y, z) coordinates of the centres for that particular region. - 'weight.txt'. Matrix (rois x rois) of SC weights. - 'tract.txt'. Matrix (rois x rois) of SC tract lengths. - 'cortical.txt'. 1 = cortical region. 0 = subcortical region. Logical. - 'hemisphere.txt'. 1 = right hemisphere region. 0 = left hemisphere region. Logical.
* sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_SC.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_SC.json*. The following CustomMetadata fields are provided:

	- 'Atlas'. Parcellation atlas used to parcel the SC.
* sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_**region_mapping**.txt	This file defines a map between sub-<participant_label>_ses-<session_label>_run-<run_label>_cortical_surface.zip and the connectivity. This file holds a vector of length equal to the number of vertices in the cortical surface. The numeric values are in the interval (0, 1, ..., n-1), where n is the number of rois in the connectivity.
* sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_region_mapping.json	This file contains the details of the parameter combinations used to generate the derived file sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_region_mapping.txt. The following CustomMetadata fields are provided: - 'Atlas'. Parcellation atlas used to parcel the cortical surface.
* sub-<participant_label>_ses-<session_label>_run-<run_label>_**cortical_surface**.zip	This compressed file contains the triangulated cortical (pial) surface mesh. The following files are provided in TVB format: - 'triangles.txt'. Matrix of (triangles x 3) indicating the combinations of indexes from the vertices.txt array forming each triangle. - 'vertices.txt'.Matrix of (NumberOfVertex x 3) indicating the coordinates (x, y, z) of each vertex in a space separated values file. - 'normals.txt'. Matrix of (NumberOfVertex x 3) indicating the normal vector (x, y, z) to each vertex in a space separated values file.
* sub-<participant_label>_ses-<session_label>_run-<run_label>_cortical_surface.json	This file contains the details of the parameter combinations used to generate the derived file sub-<participant_label>_ses-<session_label>_run-<run_label>_cortical_surface.zip.
* sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_DFA.mat	This file contains the Detrended Fluctuation Analysis (DFA) exponents for the current subject.

* sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_DFA.json	This file contains the details of the parameter combinations used to generate the derived file sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_DFA.mat. The following CustomLabel fields are provided:
	- 'MorletBank' List of log-spaced equidistant frequencies from 2 to 45 Hz.
	- 'DFAresults' Structure containing:
	  - 'group' Variable representing the atlas and the task.
	  - 'valueLabels' Labels explaining data in DFAresults.exp.
	  - 'exp' Resultant DFA exponent (parcels x fitting method (least square and bisquare) x frequencies).
	  - 'DFfunction' Values of fluctuation function calculated over the window sizes in DFAresults.winsizes.
	  - 'winsizes' Values of the window sizes.

*sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_BiS.mat	This file contains the bistability measures for the current subject.
*sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_BiS.json	This file contains the details of the parameter combinations used to generate the derived file sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_BiS.mat.

	The following CustomLabel fields are provided: - 'BiS_400ParcByFreq': Resulting bistability values for 400 parcels calculated at the frequencies in filterBank. -'filterBank': Frequencies.
*sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_pAC.mat	This file contains the phase autocorrelation measure for the current subject.
*sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_pAC.json	This file contains the details of the parameter combinations used to generate the derived file sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_pAC.mat. The following CustomLabel fields are provided: -'freq': Cut-off frequency of bands whose average is taken as to calculate the lags. -'cycles': Number of cycles for which the lag is introduced in the signal. -'PLV_values': Resulting PLV values for the bands in freq. -'time': Time in seconds for each band for which PLV_values are calculated.

5.   Overview of SEEG-BIDS

Data are shared in a BIDS-compliant derivatives structure. The folder *crosspy* i.e. the name of the pipeline we have been using and developing contains a series of folders/files, one for each analyses. All files/subfolders included here follow BIDS-convention for filenames and contain results obtained by analysing one run (run-01) of 10 minutes or resting-state (task-rest) SEEG data relative to one representative subject (sub-55). All metrics have been computed for all possible channels (univariate) and channel pairs (bivariate) and were subsequently morphed to the 100-parcel Schaefer parcellation (Schaefer 2018) and for phase lags, also to subcortical regions.

Specifically, the derivatives/crosspy folder contains:
1. *sub-55_parcel_phaselags.npy:* This file contains the resulting phase lag after morphing individual channel pairs to parcel pairs. Values are represented as a 193x193x10 matrix where the first 93 elements of the first two dimensions of matrix represent subcortical parcels, and the latter 100 elements describe the cortical parcels from the Schaefer 100 parcellation" as defined in Freesurfer.

2. *sub-55_parcel_Bistability.csv:* bistability indices, one for each parcel and for each of the 20 frequencies analysed (see section 7 for details)

3. *sub-55_parcel_CFC.npy:* Each file contains a matrix of cross-frequency coupling values between cortical parcels p_1, p_2, where p_1 (first dimension) is the parcel of the low frequency (LF) oscillation and p_2 (2nd dim.) the parcel of high frequency (HF) oscillation. The values on the diagonal represent local CFC, all others inter-areal CFC.

In the folders "CFS values" and "PAC values" all coupling values between non-epileptic contacts are given. In the "surr" folders, the surrogate values for each pair of non-epileptic contacts are given. In the "sign" folders, only those values between non-epileptic contacts remain which were found to be significant ($p < 0.01$) compared to surrogate data. In the "sign, corr" folders, only those significant values between non-epileptic contacts remain which were found to be non-spurious.

4. *sub-55_parcel_dfa.csv:* is a csv file that contains the dfa exponents for all parcels and for all frequencies (N=20) considered.

5. sub-55_parcel_static_phase_synch_spectrum.mat: contains the PLV and iPLV values for all parcel pairs and for 50 frequencies analysed.

6. *sub-55_parcel_pAC.csv:* is a .csv table holding the results of the phase autocorrelation analyses described in section 7 section. Each row contains phase autocorrelation length (in cycles) values for a range of frequencies averaged across all contacts for corresponding parcel.

6. Technical description of work performed

6.1. Materials and Methods

6.1.1. Pilot dataset

MEG dataset

A pilot dataset was created to provide an interim sample of magnetoencephalographic data for partner work packages of the TVB-Cloud consortium. The raw data were pooled from the two main projects from the UCM database and include an eyes-closed resting-state paradigm. Healthy controls (HC; n = 5), adults with Subjective Cognitive Decline (SCD; n = 5), Mild Cognitive Impairment (MCI; n = 5), and patients with Alzheimer's Disease (AD; n = 5) are included. The aim is to gain insight into how the specificities of our data map onto the requirements of partner groups in order to adjust our output to their demands. In this sense, this pilot dataset should be regarded as a growing database that will be progressively increased and updated after interaction with TVB-Cloud partners up to Month 24 when the full cohort is expected to be finally released (Deliverable 3.5: Full-cohort MEG time series in BIDS share).

All raw data from this pilot dataset were acquired between the years 2011-2016 using an Elekta Vectorview MEG system and the same acquisition protocols. The sample was recruited from the Neurology and Geriatric Departments of the University Hospital San Carlos (Madrid, Spain), from the Centre for Prevention of Cognitive Impairment (Madrid, Spain) and the Seniors Centre of Chamartín District (Madrid, Spain). All the subjects were right-handed and native Spanish speakers. A description of the data is given in **Table 6.** These data were pooled from the databases of two projects supported by the Spanish Ministry of Economy and Competitiviness (PSI2009-14415-C03-01, PSI2012-38375-C03-01).

In order to assess the general cognitive and functioning status in the original sample, a set of screening questionnaires were administered to every participant: The Mini-Mental State Examination (MMSE), the Geriatric Depression Scale - Short Form (GDS-SF), the Hachinski Ischemic Score (HIS) and the Functional Assessment Questionnaire (FAQ). After the initial screening, all subjects underwent an exhaustive neuropsychological assessment that is not included in this deliverable but that can be provided upon request (via e-mail to the co-IP Fernando Maestú: fernando.maestu@ctb.upm.es, in copy to Isabel Suárez: isabel.suarez@ctb.upm.es and Gianluca Susi: gianluca.susi@ctb.upm.es).

Cognitive concerns for Subjective Cognitive Decline (SCD) were self-reported by the participants in an interview with clinician experts and final group assignment was agreed on by multidisciplinary consensus (neuropsychologist, psychiatrists, and neurologists) following the neuropsychological evaluation. Possible confoundings of SCD such as medication, psycho-affective disorders, or other relevant medical conditions were dismissed by the clinicians. Following the recommendations made by the SCD-I-WG, all subjects were older than 60 at the onset of SCD having occurred within the last 5 years. MCI diagnosis was established according to the National Institute on Aging-Alzheimer Association (NIA-AA) criteria[4]: (1) self- or informant-reported cognitive complaint; (2) objective evidence of impairment in one or more cognitive domains; (3) preserved independence in functional abilities; and (4) not demented (McKhann et al., 2011). All patients classified as AD fulfilled NINCDS-ADRDA criteria of probable AD[5]. These criteria require patients to meet the clinical criteria for all-cause dementia along with insidious onset; clear history of worsening of cognition by report or observation; and initial and most prominent cognitive deficits that include amnestic presentation and/or deficits in language presentation, visuospatial presentation and executive function.

Table 6: Demographic information for the pilot dataset. Diagnosis: 1 = Healthy controls. 2 = Subjective Cognitive Decline. 3 = Mild Cognitive Impairment. 4 = Alzheimer's Disease.

sub-	001	004	006	007	009	010	011	013	014	015	024	025	086	087	088	134	137	139	140	141
Diagnosis	1	2	3	1	1	2	1	2	1	2	3	2	3	3	3	4	4	4	4	4
Age	67	66	69	78	72	70	69	80	70	69	81	67	75	76	78	77	63	80	80	83
Sex	M	F	M	F	M	F	M	M	M	M	F	F	M	M	M	M	M	F	F	F
MMSE	30	29	28	29	27	28	28	29	28	30	26	29	28	25	30	-	-	16	-	-

MEG data were acquired using a 306-channel Elekta Vectorview system (Elekta AB, Stockholm, Sweden) placed inside a magnetically shielded room (VacuumSchmelze GmbH, Hanau, Germany) located at the Laboratory of Cognitive and Computational Neuroscience (Madrid, Spain). MEG signals were recorded while the subjects were awake, sitting comfortably and with their eyes closed. All participants were asked to remain relaxed and stay still with their eyes closed for the duration of the recording. What participants had been doing before the eyes-closed recording varied depending on the project that they were recruited for, and is a potential source of unmodelled variance, contributing to, for example, tiredness and nervousness. However, if a subject reported feeling sleepy during the session, they were

[4] Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):270-279. doi:10.1016/j.jalz.2011.03.008

[5] McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):263-269. doi:10.1016/j.jalz.2011.03.005

given enough time to feel more awake and the recording was performed again. Prior to the recording, two electrodes were placed above and below the left eye, in a bipolar montage, to acquire electro-oculographic activity. In addition, four Head Position Indicator (HPI) coils were placed in the head of the subject, two in the forehead and two in the mastoids. The position of the 3 fiducial points, along with the HPI coils and over 200 evenly spaced points of the head shape of the subject, were acquired using a three-dimensional Fastrack digitizer (Polhemus, Colchester, Vermont). The HPI coils were fed during the whole acquisition, allowing for offline estimation of the head position. 4 minutes of resting-state activity were acquired from each subject. Data was online filtered between 0.1 and 330 Hz and digitized using a sampling rate of 1000 Hz.

6.1.2. MRI data

MEG recordings were accompanied by MRI recordings acquired within a month after MEG acquisition. The MRIs were T1-weighted, recorded in a General Electric 1.5 Tesla Magnetic Resonance scanner, using a high-resolution antenna and a homogenization PURE filter (Fast Spoiled Gradient Echo sequence, TR/TE/TI = 11.2/4.2/450 ms; flip angle 12°; 1 mm slice thickness, 256×256 matrix, and FOV 25 cm). MRIs are provided for each subject after being de-faced to account for anonymization and comply with Spanish and European Law.

6.1.3. Constraints on sharing MEG raw data

We are prevented from sharing the MEG raw data as per sharing agreement. In this pilot dataset, processed MEG data for all subjects is stored with a code-number that works as a pseudonym. The MRI data has been subjected to a defacing algorithm in order to remove facial features (while keeping the noses to allow co-registration).

SEEG dataset

The monopolar (with shared reference in the white matter far from the putative epileptic zone) local field potentials (LFPs) from brain tissue with platinum–iridium, multi-lead electrodes are recorded. Each penetrating shaft has 8 to 15 contacts, and the contacts were 2 mm long, 0.8 mm thick and had an inter-contact border-to-border distance of 1.5 mm (DIXI medical, Besancon, France). We acquired an average of 10 minutes of uninterrupted spontaneous activity with eyes closed with a 192-channel SEEG amplifier system (NIHON-KOHDEN NEUROFAX-110) at a sampling rate of 1 kHz. The electrode contacts that demonstrate non-physiological activity from analyses were excluded. We employed a novel referencing scheme for SEEG data where electrodes in grey-matter were referenced by the contacts located in the closest white-matter (CW). This referencing scheme is proven optimal for preserving phase relationship between SEEG contact data.

6.1.4. Pre-processing pipeline

MEG data

The pipeline for pre-processing the data was agreed between UH and UCM on a group a meeting held in Helsinki on May 8th, 2019, considering minor changes to adjust for each group's preferred methods. The recordings were processed offline using the spatiotemporal extension of the signal separation algorithm (tSSS) (Taulu and Simola, 2006). Parameters for the tSSS consisted of a window length of 10 seconds and a correlation threshold of 0.9. This algorithm removes the signals whose origin is estimated outside the MEG helmet while keeping intact the signals coming from inside the head. In addition, the

continuous HPI acquisition, combined with the tSSS algorithm, allowed for the continuous movement compensation. As a result, the signals used in the next steps come from a set of virtual sensors whose position remains static with respect to the head of the subject. Data from subjects whose movement during the recording was larger than 25 mm were discarded, following the recommendations of the manufacturer.

MEG data were examined using the Automatic Artifact Detection algorithm of the FieldTrip Toolbox for Matlab (Oostenveld et al., 2011) to remove ocular, muscular and jump artifacts. A MEG expert confirmed the detected artifacts. Muscular and jump artifacts were marked and segments containing them were completely discarded. On the remaining segments, a blind source separation algorithm based on second-order statistics (SOBI) was used to obtain statistically independent components. SOBI components were labeled as oculographic, cardiographic, noisy or real data. Artifact-related components were eliminated, and segments containing persistent oculographic artifacts were removed. Last, data was segmented in 4-seconds epochs of artifact-free data. Data with less than 20 epochs were discarded from the analysis, due to a low signal to noise ratio.

SEEG data

Prior to the main analysis, SEEG time series were filtered 30 Morlet filter with omega=5 and centered in frequencies ranging from 3 up to 500Hz. We excluded all 50 Hz line-noise harmonics using a band-stop equi-ripples FIR filter with 1 % of maximal band-pass ripples and 3 up to 8Hzwidth for the stop band parameters. In SEEG clinical settings, channel position are usually localized by means visual investigation of super-imposed pre-implant and post-implant imaging. In this work for assessing the position of eachrecording contact in patient's brain anatomy, we used rigid co-registration algorithm that maximizes the shared mutual information between the two images by applying an affine transform with 12degrees of freedom. Cortical pial and grey-white matter surfaces were extracted from pre-implantT1-weighted MRI volume with Freesurfer. We developed a semi-automated segmentation algorithm that uses the planned insertion coordinates (i.e. entry and target points) and the known physical dimensions of each implanted electrode (i.e. number of contacts and inter-contact distances) to identify the contact positions from post-implantation CT imaging data.

6.1.5. Source reconstruction pipeline

MEG data

The source model consisted of 2459 sources placed in a homogeneous grid of 1 cm in a Montreal Neurological Institute (MNI) template. Using the previously acquired T1 image, the MNI defined grid was transformed to subject space by an affine transformation. Each source was assigned to one of the cortical areas defined in the Schaefer 400 and the Desikan-Killiany atlases. In addition, a binary mask for the subject brain was generated using those voxels whose combined probability of WM, GM and CSF were greater than 0.5. Last, a mesh surface was generated from the previous mask using the Fieldtrip Toolbox. Finally, both the grid and the brain surface were manually realigned to the Neuromag coordinate system using the three fiducials and the head shape as guides. A lead field was calculated using a realistic single-shell head as the forward model (Nolte, 2003).

Data were filtered in the classical bands (theta, 4-8 Hz; alpha, 8-12 Hz; beta, 12-30 Hz; gamma, 30-45 Hz and broadband, 2-45 Hz) using a 2000th order FIR band-pass filter designed with a Hanning window. 2 seconds (2000 samples) of real data were used as padding in each side and discarded after the filtering.

The source reconstruction was performed using a Linearly Constrained Minimum Variance (LCMV) beamformer (Van Veen et al., 1997) for broadband. The resulting spatial filters were projected over the maximal radiation direction, getting only one filter per source. Then, the source-space time series were reconstructed by multiplying the spatial filters to the band-filtered data in sensor space.

7. Results

7.1. Local Dynamics

7.1.1. Phase autocorrelation (pAC)

Oscillatory activity has been studied for several decades in neuroscience and some methods have been proposed to quantify the amount of pure oscillatory data as well as to separate them from background activity. Here we wanted to investigate further the oscillatory nature of brain signals by quantifying the autocorrelation of phase distribution across time. In case of pure oscillatory signals, phase auto correlation would show periodic peaks while pure white-noise would rapidly decrease.

MEG data

We computed the phase autocorrelation as the phase locking value (PLV_{AC}) between the phase angles of each time series with its time-lagged versions. The phases are computed by defining $x'(t) = x(t) + iH[x(t)]$ as the analytical representation of the signal $x(t)$, where H[··] denotes the Hilbert transform, and the PLV is computed as:

$$PLV_{AC} = \frac{1}{N}\left|\sum_{n=1}^{N} e^{j(\theta(t_1)-\theta(t_2))}\right| \qquad (1)$$

where $\theta(t_1)$ and $\theta(t_2)$ are the time series' instantaneous phases at time points 1 and 2 (Lachaux et al. 1999). The PLV_{AC} values are calculated at average of cut-off frequencies for bands: Theta, alpha, Beta1, Beta2, Beta and Gamma having cut-off frequencies of [4,8], [8,12], [12,20], [20 30], [12 30] and [30 45] respectively, and with 25 time-lags are used each representing one cycle of given frequency.

References

Lachaux, J.P., et al.: Measuring phase synchrony in brain signals. Hum. Brain Mapp. **8**(4), 194–208 (1999).

SEEG data

We computed the phase autocorrelation as the PLV between the phase angles of each time series with a time-lagged version, focusing only on periods where the phase could be considered stable. Stable periods are defined similar to phase lag analyses described below (see subsequent subsection). Here, we retained only those periods where phase was stable within each single contact. Hence, for each channel and for each pair of stable periods, we quantified the phase autocorrelation as the PLV of auto-phase differences at varying shifts in time, measured as percentage of cycles of a given oscillation. Finally we asked whether the oscillations bursts were phase auto-correlated across time. We achieved so by considering just the stable periods and by measuring the phase autocorrelation in both within and between time-window pairs where phase was stable.

7.1.2. Phase lags and phase difference distributions

While the static PLV connectome yields estimates of the overall phase lags, it is a composite of possibly many distinct and dynamically variable phase lags and is hence a priori uninformative. We thus assessed the phase-lags with a time-windowed analysis so that phase-coupling between contact pairs was measured in windows of five oscillation cycles. Classical measures such as Phase locking value (PLV) is known to be biased by number of samples leading to very different estimates depending on time-window length. Here we adopted a more robust measure of phase synchrony, the Pairwise Phase Correlation (PPC). Hence time windows with PPC > 0.5 were ultimately selected for the analyses of phase lags. A high PPC threshold ensured that the chosen windows reflected strong neuronal phase coupling with a consistent phase relationship.

Given the complex variability in co-activation patterns, even a more robust time-lagged analysis might still be confounded by non-physiological or pathological (i.e. epileptic spikes) analyses. Hence we further considered only those periods where phase remained stable in both contacts in a given edge. We defined as stable those time-windows where the instantaneous frequency (i.e. the first derivative of phase time-samples) remain in between the 5th and the 95th percentiles. For all contact pairs yielding more than 30 significant, phase-stable time windows, the phases of these cPLVs were then pooled across and constituted the contact-pair-specific phase-difference distribution (PDD). To test whether these phase difference distributions were significantly different and not dominated by, spurious zero-lag derived by signal mixing, we developed a surrogate approach that would yield a similar phase synchronisation strength (PLV) with a rotated and optimally mixed signals. We then replicated the above described analyses on the surrogate data and repeated for 100 random rotation for each edge separately.

7.1.3. Long-range temporal correlations

Here we used the linear detrended fluctuation analysis (DFA) to assess the long-range temporal correlations in neuronal oscillations (Hardstone et al., 2012; Peng et al., 1994). The obtained DFA exponent, similar to the Hurst exponent, reflects the self-affinity in narrow-band amplitude fluctuations from a recorded brain region (Eke et al., 2000). To assess the DFA exponent, we followed the standard procedure (Hardstone et al., 2012).

MEG data

We obtained narrow-band amplitude time series for 26 log-spaced equidistant narrow-band frequencies from 2 to 45 Hz (Morlet m = 5, sampling rate fs = 1 kHz, central frequencies fc = [2.29, 2.57, 2.89, 3.26, 3.67, 4.12, 4.64, 5.22, 5.88, 6.61, 7.44, 8.38, 9.43, 10.61, 11.94, 13.43, 15.12, 17.01, 19.14, 21.54, 24.24, 27.28, 30.70, 34.55, 38.88, 43.75]. We used an open source toolbox named 'Neurophysiological Biomarker Toolbox (NBT)' (Peng et al. 1995) to calculate DFA exponents.

References

Eke, A., Hermán, P., Bassingthwaighte, J.B., Raymond, G.M., Percival, D.B., Cannon, M., Balla, I., and Ikrényi, C. (2000). Physiological time series: Distinguishing fractal noises from motions. Pflugers Arch. Eur. J. Physiol.

Peng et al., Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, Chaos (5), 82-87 (1995).

SEEG data

Following procedure has been applied to calculate the DFA exponents for SEEG data

1) After SEEG data preprocessing (Arnulfo et al., 2015), we obtained narrow-band amplitude time series for 20 narrow-band frequencies (Morlet m = 5, sampling rate fs = 1 kHz, central frequencies fc ranging from 2 to 225 Hz equally-spaced on log_{10} scale).

2) In each narrow-band frequency, for every SEEG contact, we computed the cumulative sum of an amplitude envelope to obtain the signal profile, *e.g.*, for a narrow-band frequency from a specific SEEG contact.

3) We then computed the fluctuation functions $F(t)$ of the signal profile with a set of window sizes T = {40 window sizes, with min and max window size being t_{min}=10 and t_{max}=60 sec respectively, and equally-spaced on log_{10} scale}. For all window sizes, neighbouring windows have an overlap of 20% (samples).

4) The DFA exponent, β, was estimated using linear regression.

References

Arnulfo, G., Hirvonen, J., Nobili, L., Palva, S., and Palva, J.M. (2015). Phase and amplitude correlations in resting-state activity in human stereotactical EEG recordings. Neuroimage 114–127.

Hardstone, R., Poil, S.S., Schiavone, G., Jansen, R., Nikulin, V. V., Mansvelder, H.D., and Linkenkaer-Hansen, K. (2012). Detrended fluctuation analysis: A scale-free view on neuronal oscillations. Front. Physiol.

Peng, C.K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H.E., and Goldberger, A.L. (1994). Mosaic organization of DNA nucleotides. Phys. Rev. E.

7.1.4. Bi-stability Index

SEEG data

The bistability index (BiS) of a narrow-band power time series (R^2) is the result of model comparison between a bimodal or unimodal fit of its probability distribution function (pdf). Here, R^2 refers to the square of narrow-band frequency amplitude time series of SEEG data. A large BiS means that the observed pdf is better fitted with bimodal, and when BiS is approaching 0, the pdf is better described as unimodal. Here, we followed the approach used in (Freyer et al., 2012; Roberts et al., 2015) to compute BiS.

1) First, to find the *pdf* of R^2, we partition observation of R^2 into 200 equal-distance bins and then count the number of observations in each bin. Next, we used the maximum likelihood estimate

(MLE) to fit a single-exponential function (the square of a Gaussian process follows an exponential *pdf*): $P_x(x) = \gamma e^{-\gamma x}$ and a bi-exponential function $P_x(x) = \delta \gamma_1 e^{-\gamma_1 x} - (1 - \delta)\gamma_2 e^{-\gamma_2 x}$ where γ_1 and γ_2 are two exponents and δ is the weighting factor.

2) Next, we compute the Bayesian information criterion (BIC) for the said single- and bi-exponential fit using $BIC = ln(n)k - 2ln(\hat{L})$, where n is the number of samples; $\hat{L}$ is the likelihood function; k is the number of parameters: 1 for single-exponential BIC_{Exp} and 3 for bi-exponential model BIC_{biE} – thus BIC imposes a penalty to model complexity (Wit et al., 2012) because the bi-exponential distribution has two more parameters (second exponents and the weight δ) than a simple exponential distribution.

3) Last, the BiS is computed as the difference between the two BIC estimates as $BiS = BIC_{Exp} - BIC_{biE}$. Because the better model will yield a small BIC value, BiS should be a large value if the bi-exponential model fit the observed power time series better.

References

Freyer, F., Roberts, J.A., Ritter, P., and Breakspear, M. (2012). A Canonical Model of Multistability and Scale-Invariance in Biological Systems. PLoS Comput. Biol.

Roberts, J.A., Boonstra, T.W., and Breakspear, M. (2015). The heavy tail of the human brain. Curr. Opin. Neurobiol.

Wit, E., van den Heuvel, E., and Romeijn, J.W. (2012). 'All models are wrong. ': An introduction to model uncertainty. Stat. Neerl.

7.2. Inter-areal dynamics

7.2.1. Functional Connectivity (FC) measures

The following functional connectivity (FC) metrics were computed and **are at the disposal of partner WPs**:

MEG data

7.2.1.1 Functional Connectomes of Phase Coupling

Phase Locking Value (PLV)

The Phase Locking Value (PLV) is a measure of the phase synchronization between two signals quantifying how their phase differences are preserved during the time course (Lachaux, Rodriguez, Martinerie, & Varela, 1999).

Here, PLV was calculated following the reformulation introduced in Bruña et al., 2017 (arXiv:1710.08037) in order to gain computational efficiency:

$$PLV_{ij} = \frac{1}{T}\left| \sum_{t=1}^{T} \dot{x}_{BP,H,i}(t) \cdot \left(\dot{x}_{BP,H,j}(t) \right)^* \right| \qquad (2)$$

where

$$\dot{x}_{BP,H,i}(t) = \frac{X_{BP,H,i}(t)}{\left| X_{BP,H,i}(t) \right|} = \frac{A_i(t) \cdot e^{-i\varphi_i(t)}}{A_i(t)} = e^{-i\varphi_i(t)} \qquad (3)$$

is the oscillatory part of the analytical band-pass filtered signal. This calculation is computationally very efficient and allows a considerable speed-up of the estimation with low memory usage.

Dynamic PLV was calculated after redefinition of the epoched time-series to define epochs of 1 second (1000 samples) with 0.5 seconds (500 samples) of overlap, thus obtaining 7 trials out of each of the former epochs. Redefined data were filtered in the classical bands (theta, 4-8 Hz; alpha, 8-12 Hz; beta, 12-30 Hz; gamma, 30-45 Hz and broadband, 2-45 Hz) using an 800th order FIR band-pass filter designed with a Hanning window. 1 second (1000 samples) of real data were used as padding in each side and discarded after the filtering. PLV was calculated following the above described method in each of the overlapping time windows to generate a dynamic connectivity matrix of dimensions [nsources x nsources x nepochs] that was then collapsed into regions of interest using the AAL atlas to render an [nrois x nrois x nepochs] matrix.

Corrected Imaginary Phase Locking Value (ciPLV)

Despite its popularity, the PLV is subject to an important limitation when used to assess functional brain connectivity: its sensitivity to volume conduction (Stam et al., 2007) and source-leakage effects. Due to the tight relation of PLV with coherency (the imaginary part of coherency discards zero-lag connectivity and thus is insensitive to volume conduction (Nolte et al., 2004)) it is possible to obtain a PLV-based measure that is insensitive to volume conduction effects. Bruña et al. (2017) defines the corrected imaginary PLV (ciPLV):

$$ciPLV_{i,j,t} = \frac{\frac{1}{T} \Im \left\{ \dot{x}_{i,t} \cdot \dot{x}_{j,t}^T \right\}}{\sqrt{1 - \left(\frac{1}{T} \Re \left\{ \dot{x}_{i,t} \cdot \dot{x}_{j,t}^T \right\} \right)^2}} \qquad (4)$$

where $\Re\{x\}$ stands for the real part of x, and $\Im\{x\}$ stands for the imaginary part of x.

Again, dynamic ciPLV was calculated after redefinition of the epoched time-series to define epochs of 1 second (1000 samples) with 0.5 seconds (500 samples) of overlap, thus obtaining 7 trials out of each of the former epochs. Redefined data were filtered in the classical bands (theta, 4-8 Hz; alpha, 8-12 Hz; beta, 12-30 Hz; gamma, 30-45 Hz and broadband, 2-45 Hz) using an 800th order FIR band-pass filter designed with a Hanning window. 1 second (1000 samples) of real data were used as padding in each side and discarded after the filtering. ciPLV was calculated following the above described method in each of the overlapping time windows to generate a dynamic connectivity matrix of dimensions [nsources x nsources x nepochs] that was then collapsed into regions of interest using the AAL atlas to render an [nrois x nrois x nepochs] matrix.

Weigthed Phase Lag Index (wPLI)

The wPLI is another metric that is insensitive to zero-lag interactions (Vinck et al, 2011):

$$wPLI = \frac{|\langle \Im(X) \rangle|}{\langle |\Im(X)| \rangle} = \frac{|\langle |\Im(X)| \, sign \, \Im(X) \rangle|}{\langle |\Im(x)| \rangle} \qquad (5)$$

where $\Im (X)$ is the imaginary component of the cross-spectrum between $\varphi_i(t)$ and $\varphi_j(t)$.

7.2.1.2. Functional Connectomes of Amplitude Coupling

Amplitude Envelope Correlation (AEC)

The envelope of oscillatory power in the frequency band of interest is obtained through the absolute value of the analytic signals related to the time series, and after being divided into n time segments of equal length Δt. Finally the Pearson correlation coefficient is computed between the envelope time series within each segment.

SEEG data

Phase synchronization (PLV/iPLV)

We also estimated inter-areal phase-phase interactions for SEEG data at individual subject level. Defining $x'(t) = x(t) + i\mathrm{H}[x(t)]$ as the analytical representation of the signal $x(t)$, where H[··] denotes the Hilbert transform, the complex PLV (cPLV) is computed as:

$$cPLV = \frac{1}{T}\sum_{t=1}^{T} \frac{x'(t)}{|x'(t)|} \frac{y'^{*}(t)}{|y'(t)|}$$

where T is the sample number of the entire signal (i.e., ~10 minutes), and * is complex conjugate. The PLV is the absolute value of complex cPLV ($PLV = |cPLV|$), and it is a scalar measure bounded between 0 and 1 indicating absence of phase and full phase synchronization, respectively.

Additionally, we used imaginary part of cPLV ($iPLV = Im(cPLV)$), a metric insensitive to zero-lag interactions caused by volume conduction for verification.

We estimated the null-hypothesis distributions of interaction metrics with surrogates that preserve the temporal autocorrelation structure of the original signals while abolishing correlations between two contacts. For each contact pair, we divided each narrow band time series into two blocks with a random time point k so that $x_1(t) = x(1\dots k)$ and $x_2(t) = x(k\dots T)$, and constructed the surrogate as $x_{surr}(t) = [x_2, x_1]$. We computed surrogate PLV across all channel pairs and assembled the surrogate interaction matrix, and its mean and standard deviation was later used in hypothesis testing.

For both PLV and iPLV connectivity, the fraction of significant edges (K) is the number of significant edges divided by the total possible edge number. Since one same white-matter contact can be used for referencing multiple cortical contacts, we rejected derivations with shared reference.

Cross-frequency coupling

SEEG data

CFS and PAC were computed between all low- (LF) and high-frequency (HF) frequency pairs at ratios of n:m (LF:HF) from 1:2 to 1:7, and for each contact-pair ca, cb of non-epileptic contacts. Frequency pairs were chosen so that the ratio of their center frequencies lay within 5% deviation of the desired integer 1:m ratio.

CFS was computed as:

$$PLV_{CFS,a,b,m} = \frac{1}{N}\left|\sum_t \exp\left[\mathrm{i}\cdot\left(m\cdot\theta_{a,LF} - \theta_{b,HF}\right)\right]\right| \tag{3}$$

where $\theta_{a,LF}$ and $\theta_{b,HF}$ are the phases of the time series of parcels. $\theta_{a,LF}$ was upsampled to match the sampling rate of the HF signal and then 'phase-accelerated' by multiplication with m. Local CFS (CFSloc) was obtained where a = b and inter-areal CFS where a ≠ b.

The strength of PAC was quantified with as:

$$PLV_{PAC,a,b} = \frac{1}{N}\left|\sum_t \exp\left[\mathrm{i}\cdot\left(\theta_{a,LF} - \theta_{b,HF,LF}^{env}\right)\right]\right| \tag{4}$$

where $\theta_{b,HF,LF}^{env}$ is the phase of the amplitude envelope of the HF signal filtered with a Morlet filter at LF, and downsampled to match the LF signal's sampling rate. Local PAC was obtained where *a = b,* inter-areal PAC where *a ≠ b.*

For both CFS and PAC, we obtained, for each frequency pair, surrogate values for each contact pair c_a, c_b by rotating $\theta_{b,HF}$ or $\theta_{b,HF,LF}^{env}$ and then calculated the means over contact pairs (PLV$_{CFS,surr\_mean}$, PLV$_{PAC,surr\_mean}$). this was done separately for inter-areal (*a ≠ b*) and local CFC (*a = b*). Connections with a ratio PLV_{meas}/PLV_{surr} of 2.42 or higher were identified as significant at p level 0.01.

In order to correct for potentially spurious observations of inter-areal PAC and CFS arising from non-sinusoidal or non-zero-mean signals, we used a novel method based on graph theory. The rationale, in brief is, that inter-areal CFC can only be spurious if the signal at f_{low} in *a* and the signal at f_{high} in *b* are also connected otherwise, namely by local CFC and inter-areal synchronization between *a* and *b*. Thus, observations of inter-areal CFC were discarded if we observed either significant local $f_{low}:f_{high}$ CFC in *p* and significant inter-areal synchronization at f_{high}, or significant local $f_{low}:f_{high}$ CFC in *q* and significant inter-areal synchronization at f_{low}.

References

Siebenhühner F, Wang SH, Palva JM, Palva S. 2016. Cross-frequency synchronization connects networks of fast and slow oscillations during visual working memory maintenance. Elife 5:e13451.

Siebenhühner F, Wang SH, Arnulfo G, Nobili L, Palva JM, Palva S. 2019. Resting-state cross-frequency coupling networks in human electrophysiological recordings. Preprint. BioRxiv: 547638.

8. Storage

8.1. MEG data

Regarding storage of the MEG-BIDS Interim Database for the Madrid Cohort (IDMC) we have set up a website located on a UCM domain, as envisaged in the Data Management Plan (DMP). The website embeds an interactive **tree viewer**, in order to provide the user a visual reference of the data structure, and the download instructions. Another separate section is dedicated to the **file manager**. Here, the user is able to download the dataset. MEG-BIDS data, as described above, are available for research purposes within the TVB-Cloud project, under the conditions specified in the DMP.
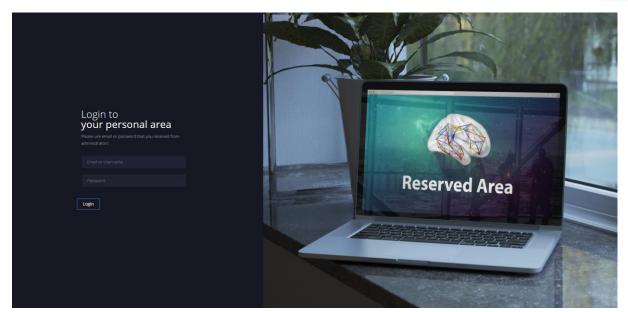
Figure 1. Interim Database for the Madrid Cohort (IDMC) Portal.

Figure 3: View of the file manager in the UCM portal for TVB-Cloud.
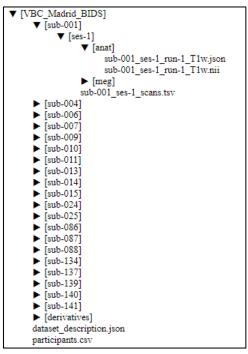

Figure 4. IDMC BIDS tree viewer in the UCM portal for TVB-Cloud.

This portal is accessible at the domain https://vbc.ucm.es. Before login, the user must request access indicating **name**, **surname**, **institution** and the TVB-Cloud **Work Package** (WP). Upon request and under IP approval, a login is created for the solicitant that allows access to the site.

Once the login has been successful, the user can download the IDMC following the steps below:

1. Download the file *BIDS_structure.rar* from the dashboard. Extract it on the user's PC to host the IDMC.
2. Go to "*VBC_Madrid_BIDS*" in the file manager and download the data for a single subject (i.e., sub-001). Extract it in the corresponding folder generated on the user's PC through Step 1 (i.e., *VBC_Madrid_BIDS/sub-001*).
3. Repeat Step 2 for each subject until folders for all subjects are completed.
4. Repeat the steps 2 and 3 for the subfolders "*VBC_Madrid_BIDS/derivatives/pipeline _preprocessing/sub-00n*", "*VBC_Madrid_BIDS/derivatives/pipeline_sources/sub-00n*" and "VBC_Madrid_BIDS/derivatives/TVB/sub-00n"
5. Please verify that the IDMC on your PC is congruent with the IDMC BIDS tree viewer, to ensure that the data extraction has been performed correctly

8.2. SEEG data

The SEEG analysis results can be downloaded after approval of request which should be sent at gabriele.arnulfo@unige.it and/or matias.palva@helsinki.fi