Assessing the depth of cognitive processing as the basis for potential user-state adaptation - Data set

The data presented here corresponds to the research study published in *Nicolae et al., (2017a)*. Other related data analysis works can be found in: *Nicolae et al., (2015a, 2015b, 2015c, 2016)*.

For any concerns and questions, feel free to use the contact information.

Introduction

The research study investigates the feasibility of quantifying how deeply the presented information is processed in the brain. Such information can potentially be used within a Brain-Computer Interface (BCI) in order to control and adapt a variety of applications, such as information seeking applications, assistance technologies for safety industrial environments and many more. In these contexts, the BCI will saliency analyze the ongoing neural activity and adapt the application according to user's current cognitive mental state. For example, in case of an increase cognitive load, the user interface can be simplified by reducing the amount of information presented, or the tasks required to the user can be diminished while the system automatically takes control over the important functions.

For this purpose, we investigate the neural correlates of different levels of cognitive processes (no, shallow and deep processing) in a visual experimental paradigm, similar to an odd-ball paradigm. While shallow processing requires a basic information processing triggered by attention (color appearance), deep processing requires the performance of a complex cognitive task related to memory, language and visual imagination processes, in addition to the attention process which is triggered in this case by color and category identification and association.

Experimental paradigm

The experiment took place in a laboratory environment. Participants were asked to stay seated, relaxed and were cued to attend the stimuli sequence, while trying to focus in the middle of the screen. However, they were allowed to freely visually explore the information presented, but to avoid any additional body movements. The visual stimulus involved a pair of images, in form of cartoon-like drawings representing a pair of objects randomly selected from a category (animals, fruits, mobility) and having one of the four colors (red, green, blue, and magenta). Firstly, the cognitive condition that needed to be performed for the current run (memory/language/visual imagination) was displayed along with the target image pair (target color and category). Further, the sequence started and participants had to distinguish first between color, and subsequently for category. When the current stimuli pair color does not match the target color (Non-Target case), no processing was requested. In case of color match (Shallow Target), the task is to perform only mental computation ('+1'). Moreover, when both the color and category of the stimuli pair matched the target color and category (Deep Target case), the task was to perform the specific cognitive process (memory, language or visual imagination) while evaluating a question-task and performing corresponding mental computations (addition). Considering the memory task, participants had to distinguish whether at least one of the images currently presented was present in the last target pair (last pair with matched target color and category). For the language task, participants had to distinguish whether the number of syllables of the object representation situated

on the left side of the image pair is greater or equal then the number of syllables of the object representation situated on the right side. As for visual imagination, participants had to imagine the object representations in real life and perform dimension comparisons, namely to check whether three times the dimension of the left object is greater or equal than the dimension of the right object. In case of affirmative answer (positive deep targets, DT+), participants have to perform a corresponding addition ('+11'), otherwise (negative deep target, DT-) nothing additional is requested (except for target counting '+1', which is required in general, for each stimulus that matches the target color). The final number counted, is entered at the end of each run. The mental counting task was integrated in order to obtain a measure of task performance and to keep the participants engaged during the experiment.

The experiment considers 15 runs, with 5 runs per condition (Memory - M, Language - L, Visual Imagination - VI), presented in this order: MMMMM - LLLLL - VIVIVIVIVI. After all 5 runs of one condition, a break of 5-10 minutes was imposed for each participant. Each run consists in 120 trials, while each condition accumulates a total of 600 trials. The trial timeline is of 2500 ms and is formed by three periods: 500 ms of pre-stimulus period with attention cue (fixation), 1250 ms of stimulus presentation and 750 ms of post-stimulus period with blank screen (relaxation).

The stimuli presentation ratio for one run is 75% NT: 12.5 % ST: 4.5% DT-: 8% DT+ with a variation of ± 2%.

For more information, please see *Nicolae et al., (2017a)* and the corresponding Supplementary Material with video examples.

Information about the data

Participants

The Electroencephalography (EEG) and behavioral data of 17 participants were recorded by members of the Neurotechnology Group at Technische Universität Berlin. No participant with neurological and/or psychiatric disorders and no pregnant woman were considered for the study. Participants were aged between 22 and 35 years old, among which thirteen were right-handed. Considering gender, eleven males took part in the study. Regarding mother tongue, thirteen participants had German mother tongue, one had English and three had other languages as native tongue. All participants had a good command of English or German, necessary for the language condition task. To increase participant's motivation, each received a remuneration of 8 €/hour. The necessary non-personal information is provided in the Excel file: "**participants_short_information.xlsx**".

The experiment was conducted in accordance with the declaration of Helsinki and approved by the ethics committee of the Department of Psychology and Ergonomics of the Technische Universität Berlin. Written informed consent was obtained from each participant: (a) to take part in the experiment and (b) to the publication of the data in anonymized form without personal information.

Considering the data analysis presented in *Nicolae et al., (2017a)*, two participant's data (participants P3 and P8) were removed from the analysis due to high artifacts and missing data segments caused by improper recordings. Specifically, for participant P3, the markers of the second run of the language condition are missing.

Behavioral data

The behavioral data comprise two indicators: an objective measure of performance represented by the final number counted (participant response number) and subjective measures regarding participants current mood (good/ok/bad) evaluated in the beginning of each run, and their personal overview referring to the difficulty of the conditions, which was acquired by scoring (0 – easy; 1 – medium; 2 – hard).

The data is available in the "**behavioral_data**" .mat file. For detailed information see General description of Matlab structs.

EEG data acquisition

The EEG data was recorded using a multichannel BrainAmp /MR Plus amplifier with 64 active electrodes ActiCap (Brain Products GmbH, Munich, Germany), arranged according to the 10-20 international system: Fp1, Fp2, AF7, AF3, AF4, AF8, F9, F7, F5, F3, F1, Fz, F2, F4, F6, F8, F10, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, EOG, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P9, P7, P5, P3, P1, Pz, P2, P4, P6, P8, P10, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, with one electrode (electrode 'T7', renamed EOG) was placed under the left eye and recorded the EOG activity, A2 and Ref (as reference electrodes placed on the left and right mastoids) and Gnd (ground electrode place at position 'AFz'). The signals were recorded at 1 kHz sampling frequency. When setting the EEG cap, the impedance was kept below $20\,k\Omega$.

The data uploaded here constitutes the raw files, namely no signal processing method was applied, except the methods described below:

- data conversion to MATLAB compatible format;
- downsampling to 100 Hz;
- re-referencing: The signals were re-referenced to left (A2) and right (Ref) mastoids channels. Therefore, the provided data comprise the rest 63 channels (62 scalp channels and one EOG).

These preliminary steps were performed with the BBCI Matlab toolbox (https://github.com/bbci/bbci_public/). For an introduction to the toolbox, refer to: https://github.com/bbci/bbci_public/blob/master/doc/index.markdown.

For details on data pre-processing and different data analysis investigations, refer to the papers in the reference list: *Nicolae et al., (2015a, 2015b, 2015c, 2016, 2017a)*.

How to load the data

We provide the behavioral and EEG data as common MATLAB (.mat) files. The data can easily be transformed into other data types, although this needs to be done manually as currently we provide no automated scripts for this. When using MATLAB, the data can be easily loaded using the *load* built-in function.

General description of Matlab structs

Behavioral data

The file "**behavioral_data**" (.mat) contains four types of information: 3 struct files (**correct_answers**, **participants_answers** and **participants_state**) and one matrix named **participants_difficulty_scores**.

- **correct_answers** {3 × 17}(1:5): contains the correct final number for each participant's experiment (17 column cells), condition (3 row cells: Memory, Language and Visual Imagination) and run (5 fields in each cell). E.g. correct_answers{2,7}(3) contains the correct final number (129) for the third run of the Language condition for the experiment with participant P7. While correct_answers{1,2}(5) contains the correct final number (118) for the fifth run of the Memory condition considering participant P2.
- **participants_answers** {3 × 17}(1:5): contains the final number counted by each participant (17 column cells) in each condition (3 row cells: Memory, Language and Visual Imagination) and run (5 fields in each cell). E.g. participants_answers{3,8}(1) provides the final number counted (152) by participant P8 in the first run of the Visual Imagination condition. While participants_answers{2,5}(5) shows the final number counted (137) by participant P5 for the fifth run of the Language condition.
- **participants_state** {3 × 17}{1 × 5}: contains information of the participants current mood at the beginning of each run. E.g. participants_state{3,9}{1,3} contains the current 'good' mood of participant P9 in the beginning of the third run of the Visual Imagination condition. Another example is participants_state{2,3}{1,4} which provides information about an 'ok' mood of participant P3 in the beginning of the fifth run of the Language condition (Remember here that second run of the Language experiment is missing, therefore the fourth run is actually the fifth).
- **participants_difficulty_scores** (3 × 17): offers information considering the rating given by each participant (17 columns) considering in general, the difficulty (scores 0, 1 or 2) of the conditions (3 rows: Memory, Language and Visual Imagination). E.g. participants_difficulty_scores(2,8) shows a medium difficulty (score 1) for the Language condition, rated by Participant P8. Another e.g.: participants_difficulty_scores(3,14) shows that participant P14 stated a high difficulty (score 2: hard) for the Visual Imagination condition.

EEG data

The "**EEG_data**" compressed 'tar.gz' file contains 51 (17x3) '.mat' files for each participant and condition, named **"ERP-Cognition-ConditionA_PX_nofilt.mat"**, where 'ConditionA' is the respective condition evaluated, namely Memory, Language or VisualImagination and 'X' is the participant's number from 1 to 17. The continuous signals from all 5 runs of each condition are contained in each file.

Each file contains a **cnt** structure containing the data, a **mrk** structure providing information about the markers and the stimuli (targets) types, and a **mnt** structure defining electrode positions, necessary for scalp visualizations. The data follows the structure defined by the BBCI Toolbox.

- **cnt**
  .clab: cell array with channel labels;

  .fs: the sampling frequency in Hz (the downsampling frequency in the present dataset);

  .yUnit: the measurement unit (µV);

  .x: continuous EEG data in form of a matrix of size [number of time points × number of channels], where the number of channels is 63.

- **mrk**

.time: the starting time point in milliseconds corresponding to the appearance of the event (stimulus). For epochs segmentation consider 500ms before this time point (pre-stimulus period) and 2000ms after (1250ms stimulus + 750ms post-stimulus);

.event: the event field contains additional information for each of the events in form of a sub-structure. Specifically, **.desc**, accessed by **mrk.event.desc**, represents information related to the type of the event, indicating for each trial, which class it belongs to: value '1' ('NT'), value '2' ('ST'), value '3' ('DT-'), value '4' ('DT+');

.y: logical matrix of size [number of classes $\times$ number of events], namely a size of 4×600. Each row of **mrk.y** corresponds to one class as specified in **mrk.className** where a value of '1' in a specific raw and column indicate that the corresponding trial ($z \in [1, 600]$) belongs to the class ($C \in$ {'NT', 'ST', 'DT-', 'DT+'}).

.className: cell array with labels for the different types of events: 'NT' (Non-Targets), 'ST' (Shallow Targets), 'DT' (Negative Deep Targets), 'DT+' (Pozitive Deep Targets).

- **mnt**

  .x: x-position of electrode for 2D plotting

  .y: y-position of electrode for 2D plotting

  .pos_3d: 3D-position of electrode for 3D plotting

  .clab: channels label information;

  .box, **.box_sz**, **.scale_box**, **.scale_box_sz**: additional information considering the electrode montage;

Selecting events from data

To access the different classes of the dataset, select the event indices in **mrk.y** that corresponds to the k-th class, by using: *find(mrk.y(k,:))* or *find(mrk.event.desc() == k)*. For example, to select the events corresponding to the first class ('NT') use *find(mrk.y(1,:))* or *find(mrk.event.desc() == 1)*. Then, *find(mrk.y(2,:))* yields the event indices for the second class ('ST'), and so on. The indices can be saved in a variable and then used to obtain the onset times of trials corresponding to the selected class. The onsets can then be used for epoching the data (e.g. with *proc_segmentation* function in the BBCI Toolbox).

Example code for events selection: *idx1 = find(mrk.y(1,:))*; *trialOnsetsClass1 = mrk.time(idx1)*. Similar functions, such as: *mrk_selectClasses* are available in the BBCI Toolbox.

Citation

The copyright is owned by the Neurotechnology Group at TU Berlin. We make the data publicly available under the Creative Commons Attribution Non-Commercial 4.0 International Public License (CC BY-NC 4.0). That means you are free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material), as long as you give appropriate credit to the author (*Nicolae et al., 2017a*). However, you may not use the material for commercial purposes.

Therefore, when you use the data, we kindly ask you to cite this data set (*Nicolae et al., 2017b*) along with the corresponding publication (*Nicolae et al., 2017a*).

Contact

For any additional information or requests, feel free to contact me at irina.nicolae@aut.pub.ro.

References

Nicolae I. E., Acqualagna L., and Blankertz B. (2015a). "Neural Indicators of the Depth of Cognitive Processing for User-Adaptive Neurotechnological Applications", *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milano, Italy, pp. 1484-1487. doi:10.1109/EMBC.2015.7318651

Nicolae I. E., Acqualagna L., and Blankertz B. (2015b) "Tapping Neural Correlates of the Depth of Cognitive processing for Improving Human Computer Interaction", *4th International Workshop on Symbiotic Interaction*, *Lecture Notes in Computer Science*, *Springer*, Vol. 9359, pp. 126-131. doi: 10.1007/978-3-319-24917-9_13

Nicolae I. E., Acqualagna L., Blankertz B. (2016). "Investigating Depth of Cognitive Processing in the Brain Dynamics of Oscillations". *Proc. of the Sixth International BCI Meeting*, California, USA, Published by Graz University of Technology (Verlag der TU Graz), p. 186. doi: 10.3217/978-3-85125-467-9-186

Nicolae I-E, Acqualagna L and Blankertz B. (2017a). "Assessing the Depth of Cognitive Processing as the Basis for Potential User-State Adaptation". *Front. Neurosci.* **11**:548, 2017a. doi: 10.3389/fnins.2017.00548

Nicolae I E, Acqualagna L and Blankertz B. (2017b). "Assessing the depth of cognitive processing as the basis for potential user-state adaptation – Data set".

Nicolae I. E., Ungureanu M., Acqualagna L., Strungaru R. and Blankertz B. (2015c). "Spectral Perturbations of the Depth of Cognitive Processing for Brain-Computer Interface Systems", *5-th edition of the International Conference on e-Health and Bioengineering (EHB 2015)*, Iasi, Romania, pp. 1-4. doi: 10.1109/EHB.2015.7391473