## Introduction

This file contains information about the EEG data that was collected from the Tata Longitudinal Study of Aging (TLSA) cohort from urban communities in Bangalore between July 2016 and July 2019, and programs used to access the data.

In addition to the TLSA cohort, two other datasets were used in one study (Murty et al., 2020, Neuroimage). This consisted of EEG data from young subjects (mainly students of IISc, used in Murty et al., 2018, JNeurosci) and another cohort of subjects less than 49 years old. These are called “VisualGamma” and “AgeProjectRound1” projects. These two datasets are not available.

## Data

EEG data were collected in 350 sessions from 279 unique subjects. For each session, the subjects were clinically diagnosed by psychiatrists (authors BN/AML in Murty et al., 2020, Neuroimage) and/or a neurologist (author MJ) as cognitively healthy, MCI or AD through clinical history and a semi-structured clinical interview.

Gamma Protocol (also called SF\_ORI since spatial frequency and orientation were varied)

1. Full Dataset: 350 sessions (279 unique subjects)
2. Discarded: 40 sessions (EEG was not even analysed for these sessions)
   1. Did not complete the experiment: 4
   2. Removed from analysis due to recording errors, poor vision etc: 17
   3. Could be replaced with a cleaner followup/baseline: 6
   4. Label (HV/MCI/AD) pending or had a discrepancy: 9
   5. Age less than 50 years: 1
   6. Assigned as a repeat even though the baseline was later discarded: 3

Remaining good sessions: 310 (257 unique subjects, HV/MCI/AD: 236/15/6)

1. No useful protocols: 11 (data was analysed but after removal of bad trials/electrodes, the session could not be used)

Remaining good sessions: 299 (247 unique subjects, HV/MCI/AD: 227/14/6)

1. Subjects with repeats: 299-247 = 52.
   1. 3 subjects had different labels in the two sessions
   2. 1 subject was MCI on both sessions

Remaining: 48 subjects who were healthy in both baseline and follow up.

SSVEP Protocol (also called TFCP which stands for TemporalFrequency-CounterPhasing)

* 1. We started with subjects who had analysable data from the gamma protocol (N=247): HV/MCI/AD: 227/14/6
  2. 8 subjects either did not do the task or the data was unusable. Left with 239 subjects (240 blocks). HV/MCI/AD: 222/12/5.
  3. Total of 31 blocks were rejected. Remaining: 209 subjects HV/MCI/AD: 197/9/3.

## Projects

The data and programs are kept in a way that additional projects can be added in a modular fashion. This data is part of four projects that are named as follows:

1. AgeProject (Murty et al., 2020, Neuroimage). How gamma varies with age in healthy subjects. N=227
2. ADGammaProject (Murty et al., 2021, eLife). How gamma varies with MCI/AD. N=227/14/6 (Healthy/MCI/AD). Two of these 14 MCIs were deemed healthy after a detailed consensus diagnosis. So the final analysis only had 12 MCIs.
3. ConsistencyProject (Kumar et al., 2022, Cerebral Cortex Communications): How gamma varies across repeat experiments. N = 48 healthy. Some other exclusion criteria were added by Santosh Kumar later, so that the final number in the paper is 40. The updated codes for this project can be found at <https://github.com/wupadrasta/TLSAEEGProjectPrograms>. Here, results are shown for 48 subjects.
4. ConnectivityProject (Kumar and Ray, 2023, European Journal OF Neuroscience): How alpha/gamma connectivity changes with age and MCI/AD.

There are other projects as well in which this dataset has been used, but the codes are not part of this repository.

## Data format

1. rawData (220 GB): files originally generated by EEG data acquisition system. All data is extracted from here. Not available in this folder.
2. SegmentedData (272 GB): Segments of data around the stimulus onset are extracted and saved from rawData. Not available in this folder.
3. cleanData (153 GB): Bad trials are removed from segmentedData using the pipeline described in Murty et al., 2020, Neuroimage. Not available in this folder.
4. decimatedData (43.8 GB): EEG data in cleanData was decimated by a factor of 10 and then saved. This dataset is freely available here: <https://osf.io/ebryn/>. The decimated data should be kept in the parent folder where the programs in this folder are kept. For example, if the programs are in X/TLSAEEGProjectPrograms, then the decimated data should be kept in X/decimatedData.
5. analyzedData: Intermediate data kept in each Project Folder.

## Programs

Programs folder contains a commonAnalysisCodes folder that contains programs that are used to identify subjects and perform standard spectral analysis. All the subfolders in this folder must be added to Matlab’s path. Also included is a file to generate decimatedData from cleanData (if you have access to cleanData), which can be done by running decimateAndSaveDataFromCleanData in informationFiles folder.

In addition, each project has a separate folder.

1. To generate analyzedData from decimatedData (if you have access to decimatedData), run runAnalyseAndSaveValuesIndividualSubject.
2. To display the results, run runDisplayData.

In addition, the following folders must be in your path for everything to work properly.

1. CommonPrograms: <https://github.com/supratimray/CommonPrograms>. The common programs used in all projects in RayLab.
2. Montages: <https://github.com/murtyvpsdinavahi/Montages> or <https://github.com/supratimray/Montages>: Contains information about layouts of different electrode caps.

These two folders could be placed anywhere and must on your Matlab path.