The data for this competition was collected from 4 human subjects. All of those subjects were epileptic patients at Harborview Hospital in Seattle, WA. At most 64 electrodes were clinically placed over frontal, parietal, temporal, and occipital cortexes. Electrocorticographic (ECoG) signals were collected during the experiment (described below). For each patient, the number of electrodes with signals in training and testing data is the same.

The *n* (*n*64) electrodes with signals that a patient has are labeled as electrode 1 to electrode *n*. The same electrode ID in two different patients does not mean that it is from electrodes in the same cortical location. Even if the electrodes are implanted at the same cortical location in two different patients, you should not assume that they are recording the neural signals of the same neuron groups. The anatomical structures of the brains of two different patients might be slightly different.

Subjects in this experiment performed a basic face and house stimulus discrimination task. They were presented with grayscale pictures of faces and houses (luminance- and contrast-matched) that were displayed in random order for 400ms each, with 400ms blank screen inter-stimulus interval (ISI) between the pictures. The 10cm-wide pictures were displayed at ~1m distance from the patients while they were seated at the bedside. There were 3 experimental runs with each patient, with 50 house pictures and 50 face pictures in each run (for a total of 300 stimuli). We called the 400ms ISI and the 400ms picture display time immediately after that a **stimulus presentation cycle**. The only **exception** is that between two consecutive runs, the inter stimulus interval of the first stimulus presentation cycle might be longer than 400ms.

The ECoG sampling frequency is 1000 Hz, therefore the inter sampling interval is 1ms. In both training and testing data, each row is a sampling point.

# More details about the experiment and the data can be found in the paper [*Spontaneous Decoding of the Timing and Content of Human Object Perception from Cortical Surface Recordings Reveals Complementary Information in the Event-Related Potential and Broadband Spectral Change*](http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004660)by *Kai J. Miller et al.*

In order to consolidate data from all four patients into a single csv file, we expand the data of patients who do not have 64 signal channels to exactly 64 signal channels, and pad the channels without signals with ***-999999***. We did this same operation on both training and testing data. The data set of the 4 patients featured in this experiment is noisier than the 7 patients used in Dr. Miller’s paper, so while the performance may be worse, it more closely reflects the practical application.

**Training data:**

For each patient, the data for the first 200 stimulus presentation cycles were put in the training data. The ending time of training data for each patient is exactly the last stimulus onset time plus 399 ms, which is when the 200th stimulus presentation cycle completes. In the training data, **the signal of each electrode from the first row to the last row is continuous**.

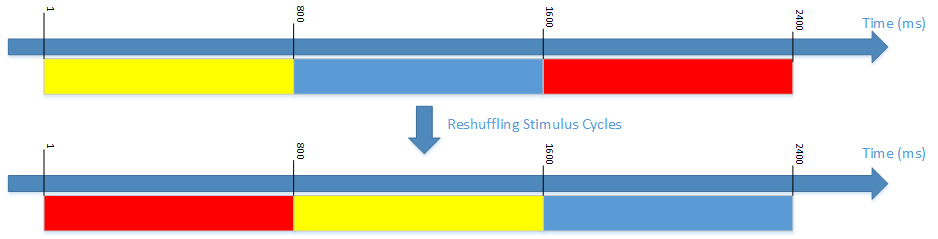
|  |  |  |  |
| --- | --- | --- | --- |
| Column Index | Variable Name | Variable Type | Description |
| 1 | PatientID | String | p1, p2, p3, and p4 |
| *2 ... 65*, | Electrode\_*i* | Float | ECoG signal channel 1…64 |
| 66 | Stimulus\_Type | Int | Integer between 0 and 101. 1-50 means house stimulus (**labeled as stimulus class 1 in this binary classification task**), 51-100 means face stimulus (**labeled as stimulus class 2 in this binary classification task**). 101 means it is inter-stimulus interval with blank image displayed. 0 also means that blank image is displayed. But 0 is not considered as inter-stimulus interval. |
| 67 | Stimulus\_ID | Int | For each patient, valued 1-200 in training data. In testing data, Stimulus\_ID is ranged 201-300, where in public testing data it is ranged between 201 and 240, and in private testing data, it is ranged between 241 and 300. |

**Testing data:**

The testing data, that your submitted web service API is called to predict, has the same data schema as the training data. In order to hold out the true labels of the test data, the true labels in column Stimulus\_Type in the testing data are all replaced with 1s when the original stimulus type is between 1 and 100. **The testing data is fixed, but not shared with participants.**

*Another difference is that the* ***stimulus presentation cycles*** *in the testing data are randomly shuffled. The data within each stimulus presentation cycle keeps its original sequence. But after shuffling, the ECoG signal between stimulus presentation cycle n and n+1 is not continuous any more.*

Here is an example to show shuffling the stimulus presentation cycles in **testing data**:



After shuffling, the original 1, 2nd, and 3rd stimulus presentation cycles become 2nd, 3rd, and 1st stimulus presentation cycles. Therefore, after shuffling, the ECoG signals surrounding the 800th sampling point are not continuous any more.

The data within each stimulus presentation cycle is not shuffled. The continuity of signals in each stimulus presentation cycle is reserved.

**How was data split into training and testing data?**

For each patient, the data up to the end of the 200th stimulus presentation cycle are put into the training data. The remainder is put in the testing data. The stimulus presentation cycles in the testing data are randomly shuffled, as described above.

**How was testing data split into public and private testing data?**

The shuffled testing data was further split into public and private testing data. For each patient, the first 40 stimulus presentation cycles (stimulus ID 201 to 240) are put into public testing data. The remaining 60 stimulus presentation cycles (stimulus ID 241 to 300) are put into private testing data.