

Background

This study consists of 4 fMRI runs – two “study” runs where people see a series of words, one at a time, and two “test” runs where they see more words, and have to indicate if each word was previously seen in the study phase or not. The study is investigating the fact that typically, words spoken out loud are later better-remembered than words read silently – this is known as the “production effect”. The goal is to determine the difference in brain activation between produced and silently read words.

Your task

Write a Python script that will turn the log files from this fMRI study (different for each subject) into files that FSL can read during fMRI analysis. This builds on the GESTASL example and you should be able to steal quite a bit of code from that one. Example output is in the intuitively-named folder.

Details you need to know

Study Phase

In the study phase, words are categorized as being in one of

three conditions: “Y”, “P”, or “S” – which refer to instructions the subject was given immediately prior to the presentation of each word:

- **P** = Produce the word (say it out loud)
- **S** = read the word Silently
- **Y** = say the word “check” no matter what word is shown (this was a control condition because directly comparing Production with Silent would show motor activation as well as memory encoding).

So for each of the two study phase runs, we need three output files for FSL (since FSL needs one file per condition). Each file should have three columns: Word Onset Time (`WORDON`), Word Duration (`WORDOFF - WORDON`), and a column of all ones. You should ignore all the rows with `NaN` in the `CONDITION` column.

Test Phase

The words in the test phase could be in one of 4 categories: any of the three above, or a new one, `FOIL`. Thus we should have 4 output files for FSL from each test phase run.

Bonus Assignment

Completing the above will suffice for a first-pass fMRI analysis. However, what will be really cool is to code the study phase items according to whether or not they were later remembered. This involves, for each word, finding it in the test phase (order of word presentation, and assignment of words to conditions, was randomized for each subject), getting the response code (`RESPONSE`) from that test phase trial, and then plugging that response into a new column in the study phase data. Then you would write out 9 log files for each study phase run, because there are three possible responses (besides `MISS` which means no response was made):

- **y** = yellow button = ? (still need to get these from Jon who wrote the program ;)
- **b** = blue button = ?
- **g** = green button = ?