Project Lambda Progress Report

Jordeen Chang, Alon Daks, Ying Luo, Lisa Ann Yu

November 12, 2015

The Paper

- A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie

About the Data

- Brief overview:
  - 20 participants recorded at a high field strength of 7 Tesla while listening to *Forrest Gump*
- Collected fMRI data for entire movie in .niigz format with additional information regarding movie scenes
- Tools used:
  - MRI scanner and pulse oximetry to conduct blood oxygen level dependent (BOLD) imaging, structural MRI imaging, and physiological assay
- Ultimate goal:
  - Provide data for others to explore auditory cognition, language and music perception, social perception, etc.

How We Chose Which Data to Use

- Files for each subject includes a lot of other information besides just the fMRI data
  - Cardiac and respiratory trace, angiographies, structural MRI data
- Three versions of the fMRI data is included:
  - Raw data, the linear alignment, and the non-linear alignment
  - Only choosing one of them: the one linearly aligned

Initial Roadblocks

- Dealing with the size of the data: 16 GB per subject, 20 subjects total
  - External hard drive to transfer data between group members
  - Limit to single subject (if time allows, will try more)
- Ambiguous direction to take the project
  - Perused scene feature data to guide possible experiments we can conduct
  - Worked on scripts to enable quick EDA iterations so we can move quickly in new directions

Current Progress

- Testing on first run for subject one
  - Data separated in 8 runs and total 20 subjects (initially work with single subject)
  - data_path.json
- Working individually on separate functions that each gather specific information
  - Plotting the standard deviations across voxels for individual subjects using Matlibplot
  - Separating data into groups based on scene details
  - Running t-tests

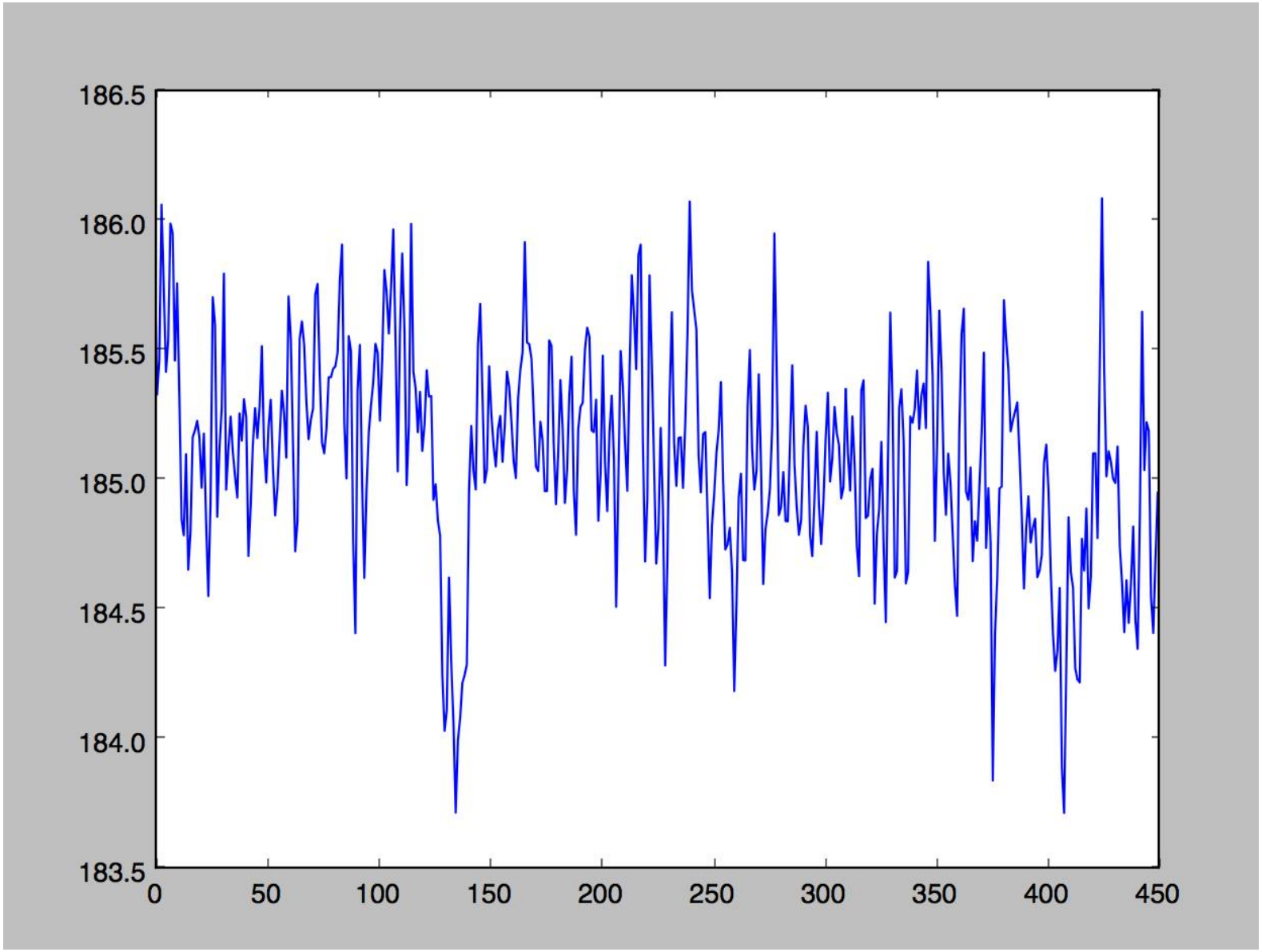

Figure 1: Plot of the standard deviations across volumes in the 4-D array for subject 1, run 1

Plan - Comparison to Original

- Deviating from the original data analysis
  - The paper used the raw data which varied largely with every subject, and processed that by standardizing among twenty of them with both linear alignments and non-linear alignments.
- Will be using the processed linearly aligned data
  - Data is already provided

Plan: Step 1 (Exploratory Data Analysis)

- Wrote functions to load the .nii files for each subject
- Calculated and plotted the standard deviations across voxels
- Using those data points to look for correlations between movie scenes and physiological responses
    - Will be cleaning out outliers before this analysis

Plan: Step 2

- Scene metadata CSV file
  - timestamp
  - brief scene description
  - day or night
  - inside or outside
- Split the images into two groups of two based on these qualities

17.0	SAVANNAH	DAY	EXT
272.0	DOCTORS OFFICE	DAY	INT
311.0	GREENBOW ALABAMA	DAY	EXT
317.0	FLASHBACK COUNTRYSIDE	DAY	EXT
343.0	GREENBOW ALABAMA	DAY	EXT

Plan: Step 3

- Perform a t-test to determine if the signal is significantly different.
- Perform a multiple comparisons test
  - To correct for the number of t-tests we will be running
- Model each of those voxels that are statistically significant to see what their time courses look like

Process - The Good

- Helpful learning how to work with the data because we have never worked with images in that type of format
  - Previously never used libraries like nibabel
  - Learn from exercises working with other fMRI data
- Team has a firm understanding of command line, python scripts, and abstraction
  - For example, not merging new functions into master unless those functions are unit tested
  - DRY

Process - The Bad

- Differing prior experiences with Github and with research
- Finding a method of communication:
  - Issues and PRs on Github are not checked immediately since main communication is Facebook
- Understanding the data - specifically in terms distinguishing the several forms of normalized data - and what conclusions we can draw is difficult
  - But we know that is beyond the scope of our project

Thank you for listening!