Predicting Imagined Meters in Musical Patterns from MEG Data

CS221 Progress Report

Zoe Robert (zrobert7), Ben Limonchik (benlimon), Aashna Shroff (aashna)

Note on Project: Due to a lack of resources at the CCRMA lab this quarter, we had to slightly pivot our experiment. Instead of predicting the direction of motion from EEG waves for use with prosthetic limbs (which would have required multiple EEG experiments at the CCRMA lab), we decided to focus on predicting imagined meters in musical patterns from MEG data. This enabled us to leverage a pre-existing database of 80 gigabytes previous collected by our mentor at CCRMA, Dr. Fujioka.

Introduction and Data Collection

Musical data is often interpreted within a metrical framework that integrates hierarchical timing information. It has been shown previously that listening to metronomes and imagining different meters such as march, waltz and hemiola (described below) modulated the resulting auditory evoked responses in the temporal lobe and motor-related brain areas such as the motor cortex, basal ganglia, and cerebellum.[1]

We wish to use MEG data of imagined meters to classify the metrical framework (i.e. march, waltz or hemiola) that the participating was imagining. The data for this project was collected in collaboration with Dr. Takako Fujioka of the Stanford CCRMA Lab. She has conducted several experiments to study the neural correlates of musical meters without the application of machine learning techniques. Our goal is to use advanced machine learning techniques to classify the imagined meters and gain quantitative insights from her findings.

Experimental Procedure: 12 healthy, right-handed musicians were studied. A constant metronome sequence of 12 unaccented clicks was played with an interval of 390 ms. The experimental conditions consisted of binary (“march”), ternary (“waltz) and “hemiola” pattern, each of which was tested in separate blocks using the identical sound sequence described above. Within each block, participants alternated between physically tapping along and imagining without movement one of the following rhythmic meters. More simply, they were to asked to tap or silently count one of the following:

March: 1 2 1 2 1 2 1 2 1 2 1 2
Waltz: 1 2 3 1 2 3 1 2 3 1 2 3
Hemiola: 1 2 3 1 2 3 1 2 1 2 1 2

Tapping coincided with the “downbeat” of each meter, i.e., the meters designated as “1”. Each participant performed approximately 190 trials. For the purposes of this project, only the portion of the data where the participant was imagining and not tapping (classifying brain waves that are coupled with some physical motion is fairly simple) was examined. Thus, the data looks as follows: 12 participants, 190 trials each, measurements over a 5 second window taken at every 0.0128 seconds, and 12,180 amplitude measurements at different spatial locations in the brain. The 12,180 measurements represent the head divided into 12,180 voxels. A visualization of the experiment is shown in Figure 1.

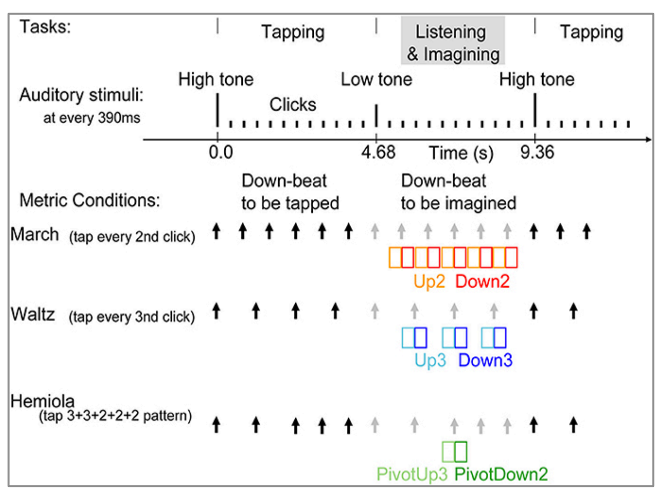

Figure 1: Experiment Stimulus and Task

[1] Fujioka, T., Ross B., Trainor L. J., Beta-Band Oscillations Represent Auditory Beat and Its Metrical Hierarchy in Perception and Imagery. Journal of Neuroscience 11 November 2015, 35 (45) 15187-15198; DOI: http://dx.doi.org/10.1523/JNEUROSCI.2397-15.2015.

Initial processing of the data involved labeling. In total, there are 9 total classifications:

'PD2_H32'		'D3_T3'
'D2_H32'	'D2_T2'	'M3_T3'
'PU3_H32'	'U2_T2'	'U3_T3
'U3_H32'		

These represent individual beats (clicks on the metronome). '_H32' is hemiola, '_T2' is March, and '_T3' is Waltz. 'D' is downbeat (i.e., all the 1's), 'U' is upbeat (i.e. the tone immediately preceding a downbeat), 'M' is middle beat, and 'P' is a pivot beats. Pivot beats appear in hemiola when switching from '123' to '12'. The data was further divided such that each individual data point is a 94 (time-intervals) by 12,180 matrix classified as one of the above labels. Overall, the data consisted 190 trials * 12 participants * 9 classifications = 20,520 data points, each with 12,180 voxels * 94 time-intervals = 1,169,280 features.

Preprocessing the data

Given the large amounts of data and the extremely large number of features it would be unreasonable to let any classification model run on the data as a whole. Therefore, the following measures were taken to limit the number of features used:

a) Any voxel with a zero-reading from the 12,180-set was removed. This resulted in 4,840 remaining cubes, since the other cubes were outside or in the periphery of the brain and were thus irrelevant.
b) Wave amplitudes measured on the peripheries of the 94-interval time-frame were found to be unreliable and greatly fluctuating from the wave amplitudes measured in between (refer to Figure 2). Therefore, the first and last 8 MEG measurements on the 94-interval time-frame were removed so only 78 time-intervals remained.

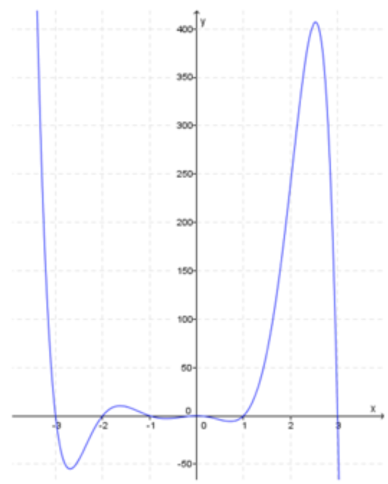

Figure 2: Typical wave graph with distortions at periphery

Consequently, each input matrix was a (78*4,840 =) 377,520 by 1 matrix, which allowed the feature set to be reduced to ~400,000 features from more than 1 million. Further reduction in features using PCA and other techniques is still required to efficiently run machine learning algorithms.

Initial Experiment and Results

For this milestone, attempts were made to classify only between 'PD2_H32' and 'D2_T2'. In particular, while analyzing MEG waves of a person imagining a binary pattern (1-2-1-2), is it possible to detect if they have been maintaining that meter or have just switched from a different metrical context (as in hemiola)?

It was unreasonable to efficiently run algorithms on ~400,000 features per data point. As a starting point, only K=100 of the 4,840 brain voxels over the 78 time-intervals were selected as features. This resulted in 7,800 remaining features. Given the large amount of data, classification was done on two participants over of 376 trials each. Therefore, the data consisted of the input X-matrix of size 756 (total trials) x 7800 (total features) and the output Y-matrix of size 756 (total trials) x 1 (0's and 1's for each label).

A **logistic regression** classifier was used for classification with **5-fold cross validation** on this modified subset of data. The results were remarkably high, with an accuracy of **98.9%**. The confusion matrix for each predicted class is shown in Figure 3.

According to prior research, the pattern of brain responses in a hemiola musical pattern significantly changes before the "pivot" point of metric transition from the ternary meter (1-2-3-1-2-3) to binary meter (1-2-1-2).[2] Therefore, the high accuracy in our results could be expected, although one might anticipate otherwise due to complex inter-temporal and spatial MEG data. It is also very likely that the large number of features in comparison to the number of observations used resulting in over-fitting noise. Therefore, a further reduction in features is necessary.

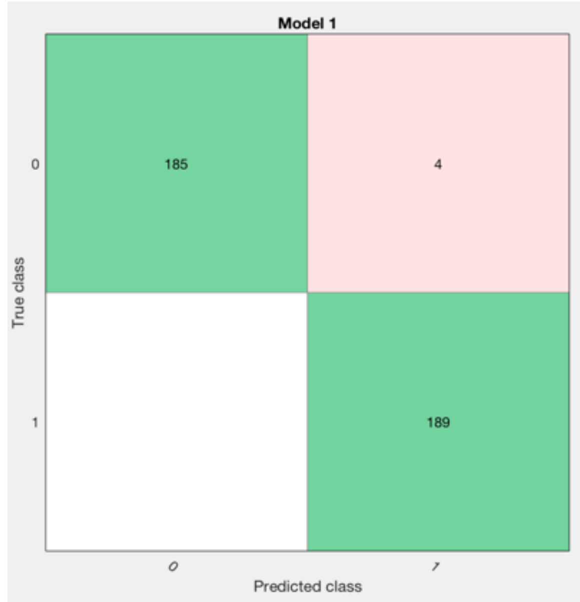

Figure 3: Confusion Matrix for Logistic Regression

Proposed Models for Future

In the future, a multiclass classification system of all 9 labels corresponding to different musical patterns will be attempted. However, a reduction in the number of features used is needed to save storage and increase comprehensibility of the data. This could be pursued in two ways:

- Feature Selection: A quick and simple filter approach to select feature subsets was applied using a univariate criterion on each feature. The two-sample *t*-test was applied and the *p*-value for each non-zero feature was compared to measure each feature's effectiveness. By plotting the CDF of the *p*-values, it was found that over 65% of the features have a p-value close to zero, and over 75% have *p*-values less than 0.05. Thus, 283,140 of the 377,520 features have strong discrimination power. Features can be sorted according to their *p*-values to select only a certain number of them. However, more analysis needs to be done to find a good estimate for number of features needed. However, such an analysis has its limitations because it does not consider interaction between the features. Linear coefficients between many features were found to be as high as 0.95; therefore, many features contain redundant information and can be removed. Alternate techniques like sequential feature selection can be pursued.

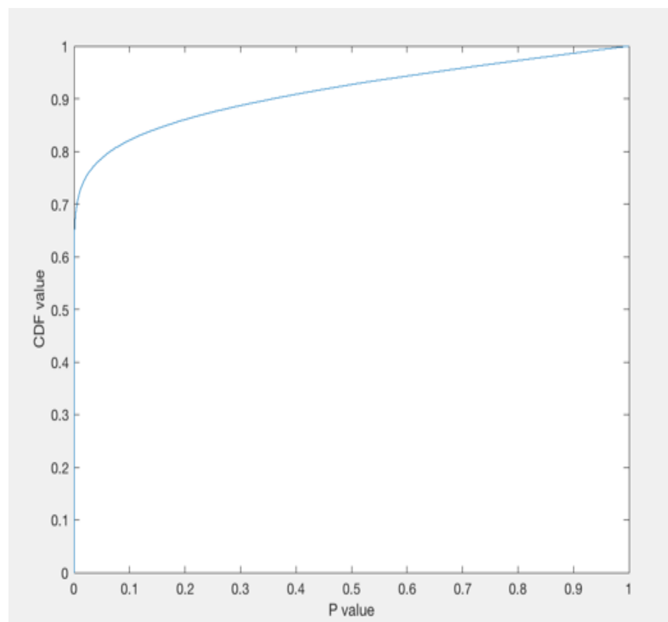

Figure 4: Empirical CDF of p-values

- Feature Transformation: Principal Component Analysis can be run to reduce the dimensionality of the data. This can be used to learn inter-spatial information about the regions of the brain that are most relevant for classification, as well as the inter-temporal information about time-intervals where most differences are observed.

After having reduced the dimensionality, a classification on a sample set of more than two participants will be pursued to avoid over-fitting and ensure generalization for more people.

[2] Fujioka, T., Fidali B. C., Ross B. Neural correlates of intentional switching from ternary to binary meter in a musical hemiola pattern. Front. Psychol., 12 November 2014, http://dx.doi.org/10.3389/fpsyg.2014.01257.