MicroMotion: Prediction of head movement based past head movement and noise

Report by: Brian Wilson

1. Project Statement

The project was based around data collected during a 2012 study conducted at the University of Oslo and resulted in a paper, Jensenius et al., "The Musical Influence on People's Micromotion when Standing Still in Groups", Proceedings of the 14th Sound and Music Computing Conference (2017). The study is described as follows:

“It is commonly assumed that listening to musical sound, and particularly dance music with a clear pulse, makes us move. However, most empirical studies of music-induced motion have mainly focused on voluntary and fairly large-scale movement. This dataset was collected as part of a study which aimed to investigate the effects of music stimuli on movement when participants try to remain at rest. We collected data through optical motion capture from groups of people instructed to stand as still as possible with and without music stimuli. We then looked at the differences in movement between conditions.” [1]

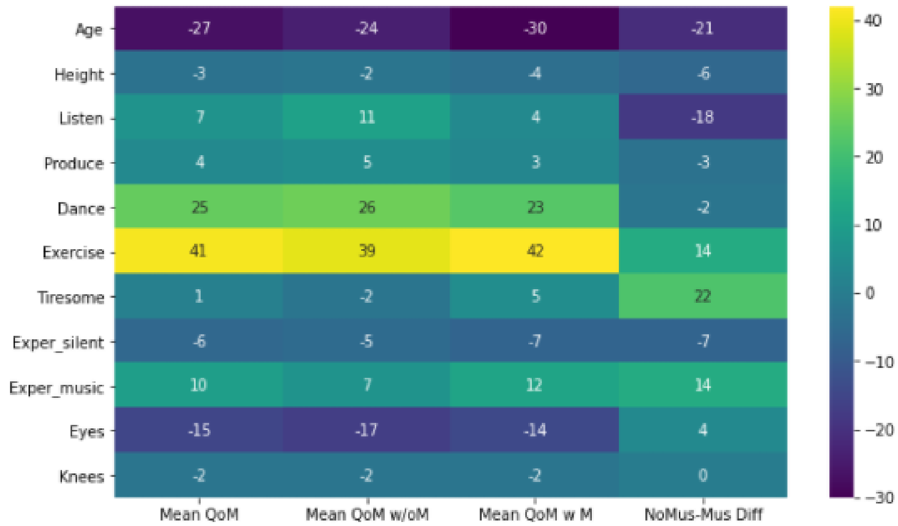
Using the data collected during the study we are looking to predict future head movement based upon silence/music at the preceding time as well as movement up until that point. If we find that there is predictive power in this previous movement then it could point to a stronger likelihood of usefulness for the field of musicology. By determining what factors appear to most be affecting unconscious movement, we could further unlock how they could affect unconscious thought. We will specifically look to predict the sum of movement over the next 5 seconds, 10 seconds and 20 seconds into the future.

2. Datasets

The collected data is represented in three categories: Motion, Music, Demographics.\

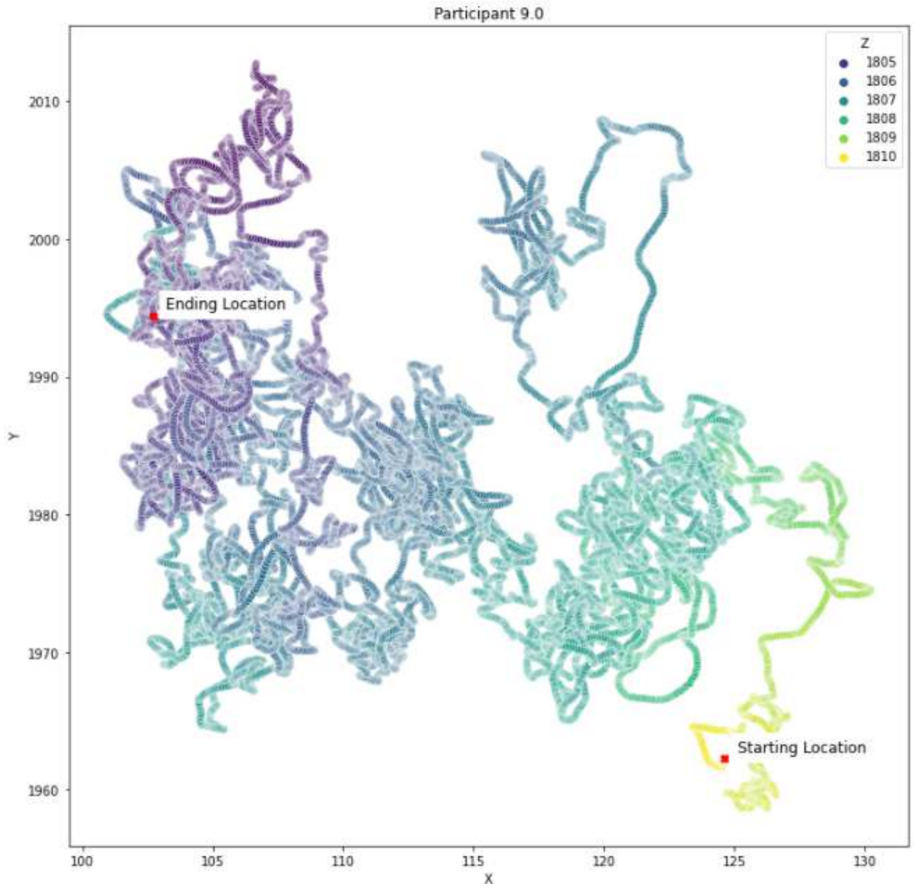
- Demographics were collected via a survey given after the experiment. For the modeling portion we removed the columns that included ex-post data to prevent the model from receiving “future” information.
- The demographics data also included quantity of motion (QoM) data for each participant for portions of the experiment with and without musical stimuli. We ran Pearson correlation that was plotted to a Heatmap to gather a general idea of what features most correlated with motion. No feature showed a correlation over 0.50 , most were much less. Exercise was the highest positive correlation at approx. 0.40 for all types of motion, while Age had the highest negative correlation with approx.. -0.28. The correlation for each relation ship is shown in figure 1 below.

Figure 1



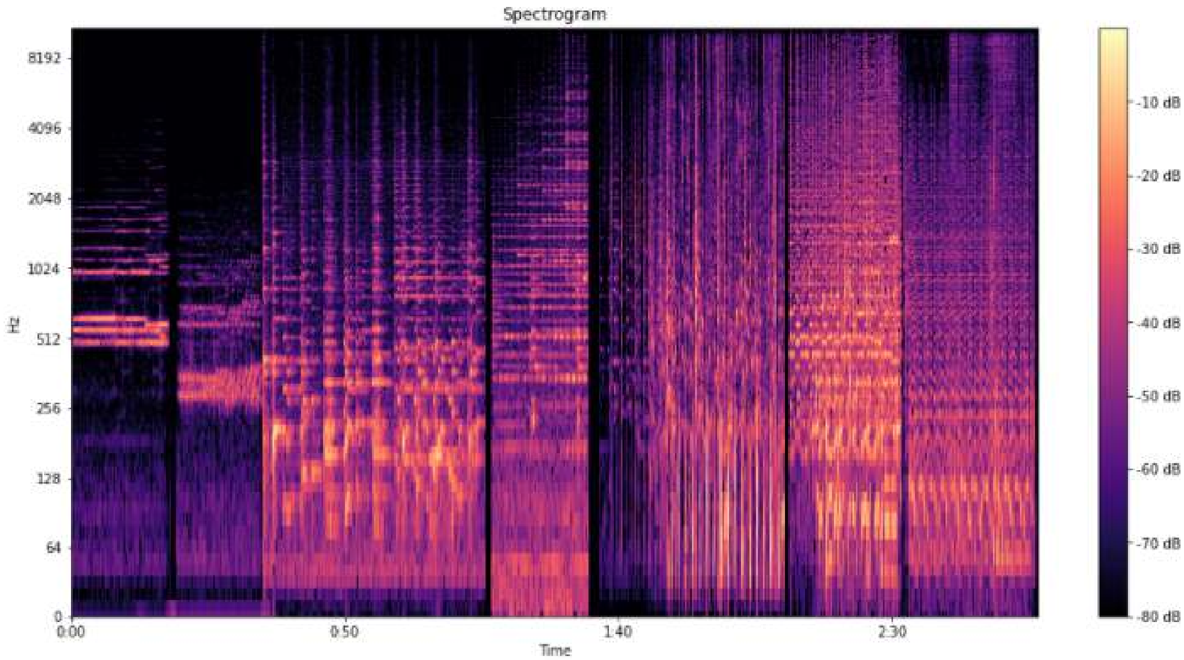
- Motion was collected as follows:
  - “The instantaneous position of a reflective marker placed on the head of each participant was recorded using a Qualisys infrared motion capture system (13 Oqus 300/500 cameras) running at 100 Hz. The data were recorded in 8 groups of 12-17 participants at a time. Participants were asked to stand as still as possible for 6 minutes, starting with 3 minutes in silence and followed by 3 minutes with music. Participants were aware that music would start after 3 minutes, and were free to choose their standing posture. The distribution of participants in the recording space was standardized across trials with marks on the floor indicating the approximate feet position. The motion capture system was triggered and stopped automatically with the stimuli playback system, thus all recordings are exactly 6 minutes long.”[1]
  - An example of the collected data is shown in Figure 2 below. The visual represents a view looking down onto the top of a participants head with the following corresponding planes
    - X – left to right, or side to side
    - Y – front to back
    - Z – height represented by color with lighter being higher and darker being lower

Figure 2

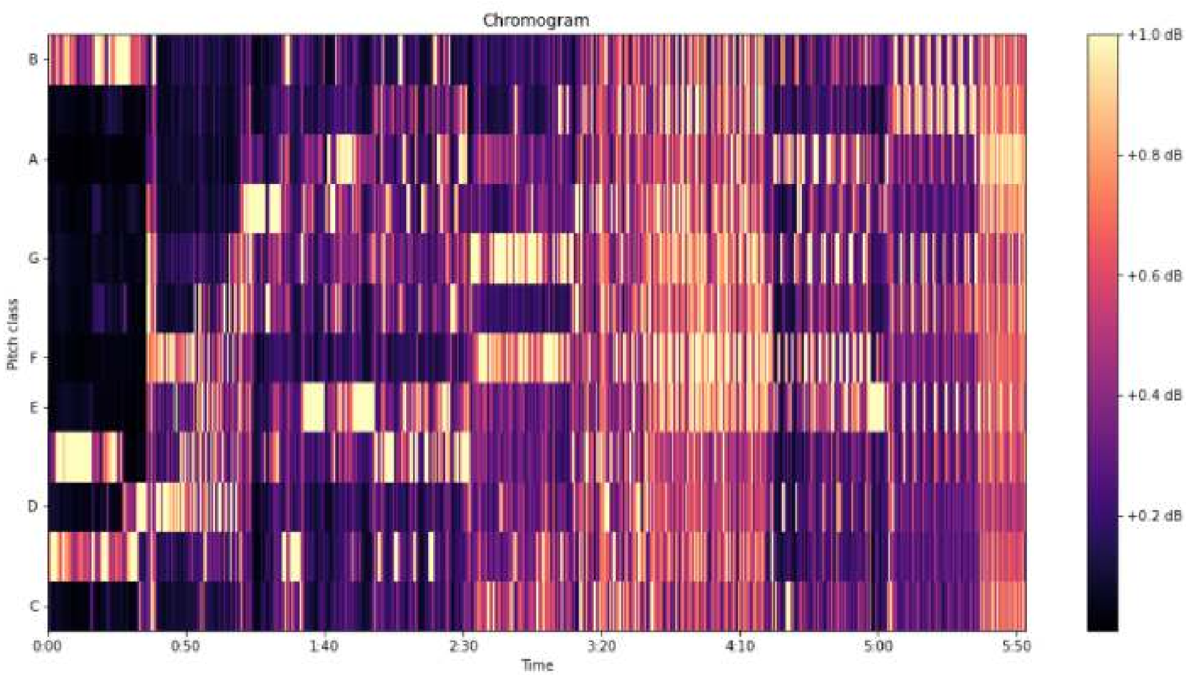


- Music used for stimuli during the later 3 minutes was the following[1]:
  - Lento (#3) from György Ligeti Ten Pieces for Wind Quintet (20s)
  - Allegro con delicatezza (#8) from György Ligeti Ten Pieces for Wind Quintet (15s)
  - Adagio from Joaquin Rodrigo's Concierto de Aranjuez (40s)
  - Winter movement from Vivaldi's The Four Seasons (20s)
  - Left & Right by D'Angelo, featuring Method Man & Redman (35s)
  - Marcando la distancia by Manolito y su trabuco(20s)
  - Cubic by 808 State (30s)
- For purposes of model building we converted the model into a chromogram using the python Librosa package. A chromogram is a representation of the strength of musical notes (12) over time periods where a spectrogram represents frequency in hertz. The following spectrogram and chromogram represents the 3 minute portion of the experiment where the music stimuli was present.

Figure 3/4



Spectrogram
8192
4096
2048
1024
512
256
128
64
0
Hz
0:00
0:50
1:40
2:30
Time
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB



Chromogram
B
A
G
F
E
D
C
Pitch class
0:00
0:50
1:40
2:30
3:20
4:10
5:00
5:50
Time
+1.0 dB
+0.8 dB
+0.6 dB
+0.4 dB
+0.2 dB

3. Cleaning and Processing

a. Demographics

i. The data was relatively clean with most columns being numeric type except for Group and Sex. The group column was used to create a subject ID column ('SID') column that retained the individual group divisions while a Participant ID column ('PID") was created to give a global designation. There were 17 participants who have no value for 'Eyes open?' and 'Locked knees'. This happened for all participants in the P group. We assigned 0.5, which represents both, for both of these columns. There was also one participant who didn't give an experience of motion answer, so we used a mean of the column to fill.

b. Motion

i. Motion data was provided for each participant, we ran the same cleaning process on each prior to combining them into a single dataset. The stationary reference points were removed from each set and the X, Y, Z data were remand as simply "X, "Y", Z". Some new features were added to each set as a basis for predicting as well as for prediction targets:

1. disp_start - The distance from the starting location at each point (not a sum)
2. scale_mean - The position scaled by the mean causing the mean for all samples to be zero
3. disp - absololute value of displacement per step
4. disp_total - sum of disp from start
5. step_eucl - Euclidean distance moved in time step
6. total_eucl - Total Euclidean distance move since start
7. target_10 - Target for prediction, 10s ahead
8. target_5 - Target for prediction, 5s ahead
9. target_20 - Target for prediction, 20s ahead
10. _s_avg - Average of noted variable for last amount of seconds
11. _s_min - Min of noted variable for last amount of seconds
12. _s_max - Max of noted variable for last amount of seconds

c. Music

i. As previously mentioned the music was pulled in based on it's note values for the 12 common notes C, C#, D, D#, E, F, F#, G, G#, A, A#, B. Features were created for each note.

ii. note + '_5sec_avg' – mean of note strength for previous 5 seconds

iii. note + '_2.5sec_avg' – mean of note strength for previous 2.5 seconds

iv. note + '_1sec_avg' – mean of note strength for previous 1 second

v. note + '_10sec_max' – max of note strength for previous 10 seconds

vi. note + '_5sec_max' – max of note strength for previous 5 seconds

vii. note + '_1sec_max' – max of note strength for previous 5 seconds

viii. note + '_10sec_min' – min of note strength for previous 5 seconds

ix. note + '_5sec_min' – min of note strength for previous 5 seconds

x. note + '_1sec_min' – min of note strength for previous 5 seconds

xi. note + '_10sec_diff' – difference between max and min of note strength for previous 10 seconds

xii. note + '_5sec_diff' – difference between max and min of note strength for previous 5 seconds

xiii. note + '_1sec_diff' – difference between max and min of note strength for previous 1 second

d. Resampling

i. The music data was brought in at 22 Hz or 22000 cycle per second. The motion data was captured at 100 readings per second. In order bring them into a single dataframe as well as with thought to memory/computation restraints, the music and motion data were both sampled to one reading every 2 tenths of a second. The mean method was used during resampling. This resulted in just under 1800 rows of data for each participant and roughly 130,000 rows of data in total.

4. Modeling

a. PCA / ElasticNet / Random Forest

i. Features were run through StandardScaler ensure they were not affected by difference in units, types of measures

ii. The musical notes explain the most amount of variance in the dataset, though this doesn't point to them being the most important features for prediction.

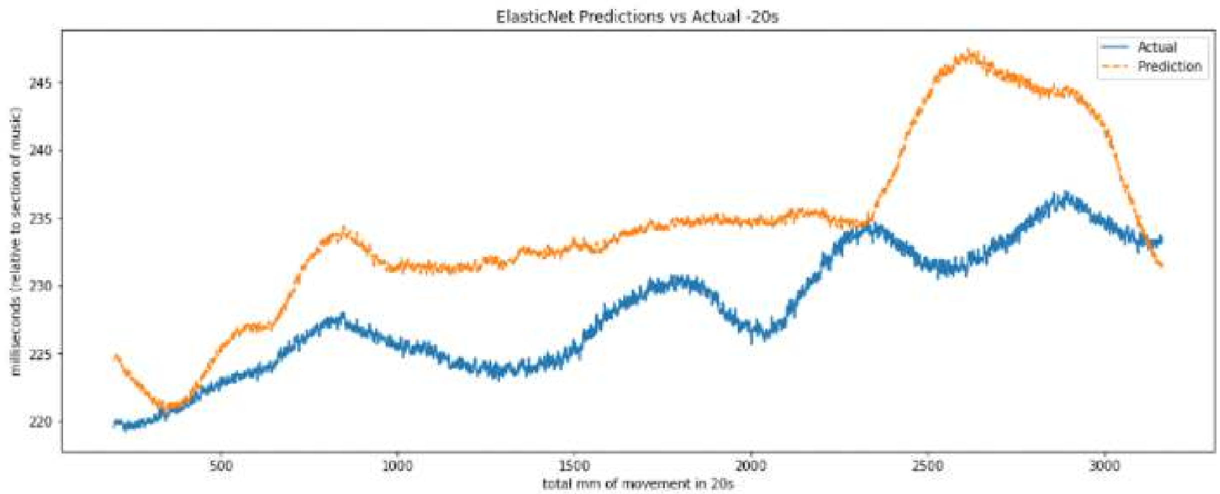
iii. We ran through a number of options including PCA with 10 and 50 components using both ElasticNet and RandomForest out of the box. We found that the use of all features performed the best with an MAE of 7.264 mm and an R^2 score of 0.47. ElasticNet performed slightly better than the RandomForest on MAE and R^2, but on inspection of the plotted predictions vs actual values RandomForest appeared to better follow the trends.

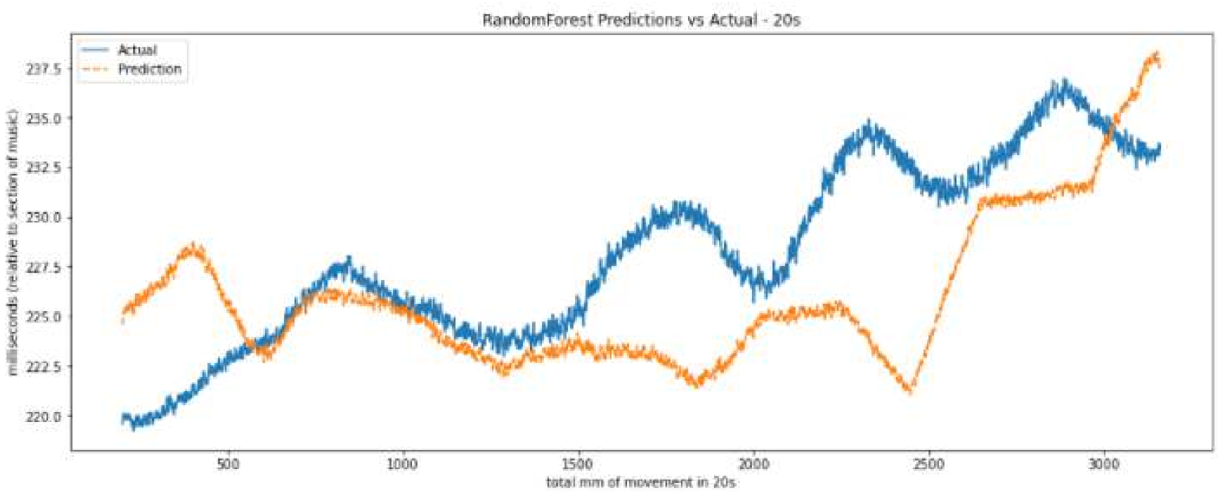
iv. We then moved onto comparison of our target variables: 5 seconds, 10 seconds, and 20 seconds. 20 seconds had the best results with MAE of 19.97 and R^2 of 0.56. This seema worse than the target 5 score above, but the R^2 score is better and the MAPE is actually less as it is a smaller percentage of the mean actual value. MAPE for target_20 : 8.4% for target_10 : 10.0% target_05 : 15.13%

	variance_explained
C	0.529037
C_sharp	0.075885
D	0.030363
D_sharp	0.028624
E	0.025924
F	0.019877
F_sharp	0.015268
G	0.014266
G_sharp	0.013044
A	0.012326
A_sharp	0.010870
B	0.010555
millisecond	0.010263
C_5sec_avg	0.009859
C_2.5sec_avg	0.008946

	Coefficents
Dance	2.436807
Z_disp	1.779447
Z_5s_max	1.712204
Y_10s_max	1.554104
X_5s_max	1.393609
Y_20s_max	1.363275
X_20s_max	1.269266
Z_20s_max	1.244244
step_eucl_5s_max	1.181014
Listen	1.138900
Y_5s_max	1.011529
Z_10s_max	0.919841
Y_10s_avg	0.744301
Tiresome	0.744021
Y_5s_avg	0.714485

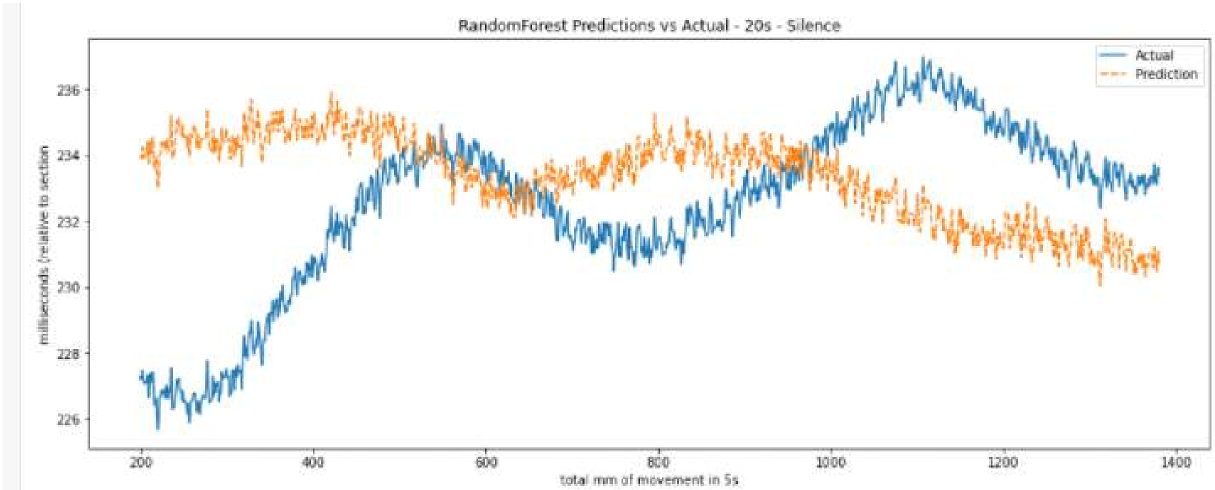
	Importances
Z_20s_max	0.729567
Height	0.071403
total_eucl	0.047080
X_20s_max	0.026838
X_scale_mean	0.016192
Z_disp_start	0.016129
X_disp_total	0.013285
X_20s_min	0.011198
Age	0.008449
Z_disp_total	0.007004
Y_disp_total	0.005020
X_10s_max	0.004360
X_10s_min	0.003770
Z_scale_mean	0.003248
step_eucl_10s_min	0.002923

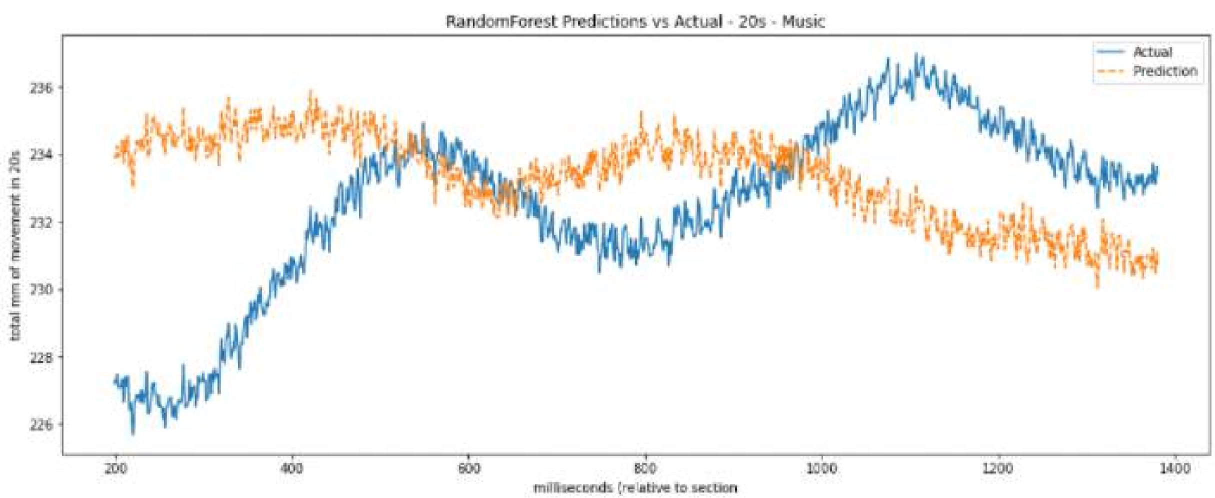


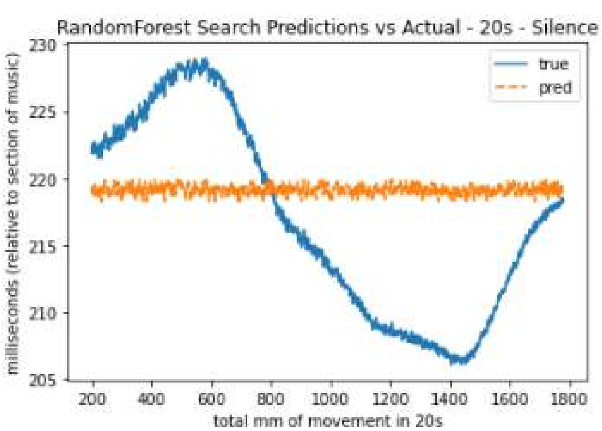


b. Music vs. Silence
   i. While we weren't finding great predictive power looking across the whole run of 6 minutes we looked at splitting the time into silence and music sections. We would run individual train and test sets in each grouping.
   ii. First we made a comparison to a mean strategy dummy variable which performed considerably worse than either our ElasticNet or RandomForest at MAE of approx. 35 for total section, just silence and just music. The R^2 was just below zero.
   iii. ElasticNet delivered a MAPE on silence of about 10% and on music of just over 7%. RandomForest delivered a MAPE of 6.6% on silence and just over 7% on music.

iv. We have so far been using participants 20 to 25 for our modeling as a sample. We will expand further as we start our hyperparameter tuning.



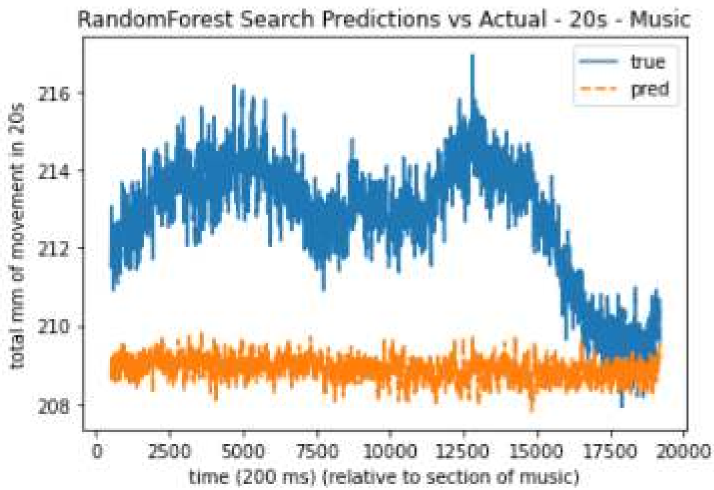




c. Randomized Search Random Forest

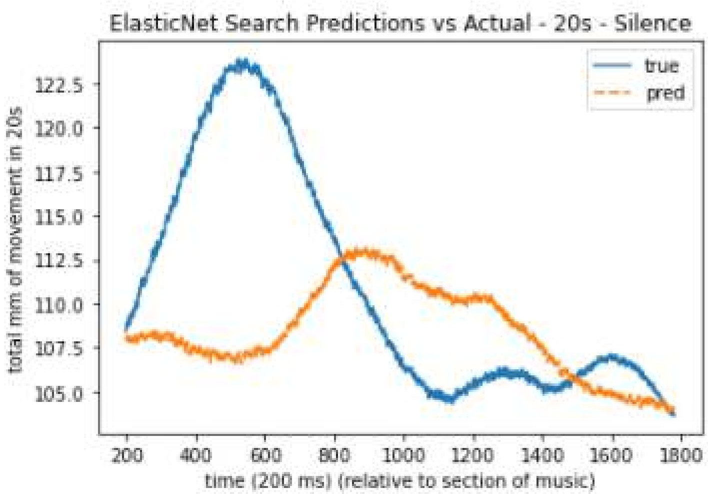
i. Results from a Randomized Search of Random Forest features for the target_20 variable during silence results in the following parameters:

1. 'min_samples_split': 0.4,
2. 'min_samples_leaf': 0.1,
3. 'max_leaf_nodes': 6,
4. 'max_depth': 5,
5. 'criterion': 'mse'



ii. Results from a Randomized Search of Random Forest features for the target_20 variable during music results in the following parameters:

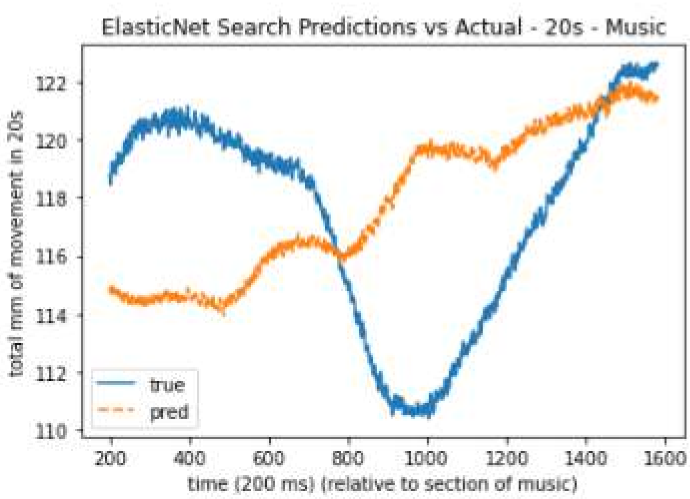
1. 'min_samples_split': 0.2,
2. 'min_samples_leaf': 0.1,
3. 'max_leaf_nodes': 4,
4. 'max_depth': 5,
5. 'criterion': 'mse'



d. Randomized Search ElasticNet

i. Results from a Randomized Search of ElasticNet features for the target_20 variable during silence results in the following parameters:

1. 'l1_ratio': 0.7,
2. 'alpha': 0.8



ii. Results from a Randomized Search of ElasticNet features for the target_20 variable during silence results in the following parameters:

1. 'l1_ratio': 0.7,
2. 'alpha': 0.8

5. Conclusions

a. Predictive or not

Based upon the results that were found the predictive power of previous motion and musical stimuli is inconclusive. The models do consistently score above a mean strategy dummy indicator, but don't do well with matching the actual trends in the data. Additional algorithms could be used with potentially better results, but there is nothing in the current results that points towards a definitive answer.

b. Future Possibilities

Different lines of inquery are still available for study such as:

- Men vs. women
- Types of music (segments) broken out
- Training on data from 2/3 of participants fully and testing on remaining
  - Together, silence separate, music sparate

[1] MICRO Motion capture data from groups of participants standing still to auditory stimuli (2012) v1.0 (physionet.org)