

# A Comprehensive Feature and Data Mining Study on Musician Memory Processing Using EEG Signals

Kin Ming Kam<sup>1</sup>(✉), James Schaeffer<sup>2</sup>, Shouyi Wang<sup>1</sup>, and Heekyeong Park<sup>2</sup>

<sup>1</sup> Department of Industrial, Manufacturing and Systems Engineering,  
University of Texas at Arlington, Arlington, TX 76019, USA  
kinming.kam@mavs.uta.edu, shouyiw@uta.edu

<sup>2</sup> Department of Psychology, University of Texas at Arlington,  
Arlington, TX 76019, USA  
james.schaeffer@mavs.uta.edu, hkpark@uta.edu

**Abstract.** There has been much interest in the beneficial effects of musical training on cognition. Previous studies have indicated that musical training was related to better working memory and that these behavioral differences were associated with differences in neural activity in the brain. However, it was not clear whether musical training impacts memory in general, beyond working memory. By recruiting professional musicians with extensive training, we investigated if musical training has a broad impact on memory with corresponding electroencephalography (EEG) signal changes, by using working memory and long-term memory tasks with verbal and pictorial items. Behaviorally, musicians outperformed on both working memory and long-term memory tasks. A comprehensive EEG pattern study has been performed, including various univariate and multivariate features, time-frequency (wavelet) analysis, power-spectra analysis, and deterministic chaotic theory. The advanced feature selection approaches have also been employed to select the most discriminative EEG and brain activation features between musicians and non-musicians. High classification accuracy (more than 95%) in memory judgments was achieved using Proximal Support Vector Machine (PSVM). For working memory, it showed significant differences between musicians versus non-musicians during the delay period. For long-term memory, significant differences on EEG patterns between groups were found both in the pre-stimulus period and the post-stimulus period on recognition. These results indicate that musicians memorial advantage occurs in both working memory and long-term memory and that the developed computational framework using advanced data mining techniques can be successfully applied to classify complex human cognition with high time resolution.

## 1 Introduction

There has been much interest in the beneficial effects of musical training on cognition. Previous studies have indicated that musical training was related to better working memory and that these behavioral differences were associated

with differences in neural activity in the brain [1]. However, it was not clear whether musical training impacts memory in general, beyond working memory. By recruiting professional musicians with extensive training, we investigated if musical training has a broad impact on memory with corresponding electroencephalography (EEG) signal changes, by using working memory and long-term memory tasks with verbal and pictorial items. Behaviorally, musicians outperformed on both working memory and long-term memory tasks. A comprehensive EEG pattern study has been performed, including various univariate and multivariate features, time-frequency (wavelet) analysis, power-spectra analysis, and deterministic chaotic theory. The advanced feature selection approaches have also been employed to select the most discriminative EEG and brain activation features between musicians and non-musicians [2]. High classification accuracy (more than 95 %) in memory judgments was achieved using Proximal Support Vector Machine (PSVM) [3]. For working memory, it showed significant differences between musicians versus non-musicians during the delay period. For long-term memory, significant differences on EEG patterns between groups were found both in the pre-stimulus period and the post-stimulus period on recognition. These results indicate that musicians memorial advantage occurs in both working memory and long-term memory and that the developed computational framework using advanced data mining techniques can be successfully applied to classify complex human cognition with high time resolution.

## 2 Methodology

### 2.1 Data Acquisition and Experimental Settings

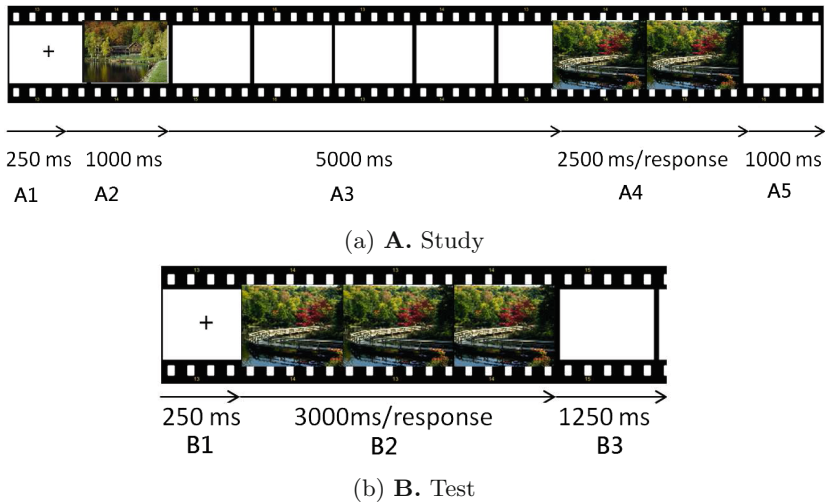
*Participants.* Initially, 36 musicians and non-musicians participated into the experiments. In those 36 participants, some of them were excluded based on behavioral observation and outlier analysis. If participants are failed to follow the instruction, they will be excluded. Two of them fell into this category and were excluded. Cook's D values of both short-term and long-term memory tests were calculated to identify outliers. Subsequently, four of them were excluded due to having negative Cook's D values on the long-term memory test. One participant achieved higher than 3 standard deviations so he was also excluded from the data. Finally, 29 subjects were remained for analysis. We had 14 professional musicians who have over 10 years of experience. Five of them were female. The average of experience is 22.9 years. We also had 15 participants without any musical training. They were marked as "non-musicians". Among them, eight were female. Informed consent was obtained from all participants in accordance with the experimental protocol approved by the University of Texas Institutional Review Board.

*Design of the Experiments.* The whole experiment was separated into two parts: 1. a study session Participants completed a study session followed by a test session involving words and pictures as stimuli. Stimuli were presented visually on a computer and all responses were made using the keyboard. During the study

session, participants were presented with pairs of stimuli, one at a time. Each study trial began with a fixation cross (250 ms), the first stimulus (1000 ms), a blank screen (5000 ms), the second stimulus (2500 ms or until a response), and finally a blank screen (1000 ms). Upon presentation of the second stimulus, participants made a judgment of whether the second stimulus was the same as the first (Fig. 1a).

A few minutes following the study session, participants memory was tested. During this test session, stimuli presented during study were presented again along with new stimuli that had not been studied. Further, we only tested participants memory on stimuli that had only been presented once. Therefore, only stimuli presented on trials that were different during the study session (i.e. trials on which the second stimulus was different from the first) were presented during test. Each test trial began with a fixation (250 ms), followed by a stimulus (3000 ms or until a response), and then a blank screen (1250 ms). Upon presentation of the stimulus, participants made a memory judgment which included a rating of how confident they were in their memory (Fig. 1b). They were allowed to make three responses: *remember with low confidence*, *remember with high confidence*, or *new*.

Word and picture stimuli were blocked for both study and test phases, such that each participant was presented with a block of word trials followed by a



**Fig. 1.** Schematic of experimental paradigm. (A1 to A5) During study period, participants were asked to judge whether the second stimulus matched the first. (B1 to B3) During test period, participants made memory judgments to stimuli while rating their confidence. Low represents remember with low confidence, High represents remember with high confidence, and New represents a judgment where participants thought the stimulus was not studied.

block of picture trials (or vice versa). Whether or not participants were presented with words or pictures first was randomly determined for each participant.

*Types of Stimuli.* Participants were presented with pictures of complex scenes and words. During the study session, participants completed 96 trials of pictures (32 same, 64 different) and 96 trials of words (32 same, 64 different). Given that each trial contained two stimulus presentations, participants studied a total of 128 pictures and 128 words from different trials. These 248 studied stimuli were then used to test long-term memory during the test session. During the long-term memory task, participants completed 192 trials of pictures (128 studied, 64 new) and 192 trials of words (128 studied, 64 new).

*EEG Data.* EEG data were collected during both study and test sessions using the Brain Vision ActiChamp 32 channel system and recorded using the Pycorder software. Electrode positions followed the 10–20 system and included Fz, Cz, Pz, Oz, Fp1, Fp2, F3, F4, F7, F8, Fc1, Fc2, Fc5, Fc6, Ft9, Ft10, T7, T8, C3, C4, Cp1, Cp2, Cp5, Cp6, Tp9, Tp10, P3, P4, P7, P8, O1, and O2 according to standard 10/20 system. During recording, data were sampled at 1000 Hz and filtered between .01 and 100 Hz. Offline, data were high-pass filtered with a 0.1 Hz Butterworth filter, downsampled to 256 Hz, and referenced to the average of the mastoids (TP9 and TP10). Post-stimulus ERPs with a 1000 ms duration were extracted and were baseline-corrected with respect to a 200 ms prestimulus baseline. Visual inspection was then used to remove epochs that contained eye blinks and movement artifacts using a recently developed automatic ICA-based algorithm, called ADJUST [5].

## 2.2 Spatiotemporal Pattern Based Artifacts Removal

Brain signals often contain significant artifacts that lead to major problems in signal analysis, when the activity due to artifacts has a higher amplitude than the one due to neural sources. The common sources of artifacts include eye movements, muscle contractions, electric devices interference [4]. Independent Component Analysis (ICA) has been successfully applied for artifacts removal in many studies. The basic idea is to decompose the brain data into independent components, determine the artifacted components using pattern and source localization analysis, and reconstruct the brain signals by excluding those artifacted components. However, linking components to artifact sources (e.g., eye blinking, muscle movements) remains largely user-dependent. In this study, we employed ADJUST for signal artifact removal. ADJUST applies stereotyped artifact-specific spatial and temporal features to identify independent components of artifacts automatically. These artifacts can be removed from the data without affecting the activity of neural sources [5]. The data analysis in the following is based on the ‘cleaned’ data after artifact removal.

### 2.3 Signal Feature Extraction

We extensively investigated features from the collected physiological signals. Four groups of feature extraction techniques were employed to capture signal characteristics that may be relevant to assess memory workload. They were signal power, statistical, morphological, and wavelet features [6]. For a data epoch with  $n$  channels, we first extracted features from signals at each channel, and then concatenated the features of all the  $n$  channels to construct the feature vector of the data epoch. The feature extraction of four groups of signal features are listed in Fig. 2 [7] (Tables 1 and 3).

**Table 1.** Frequency ranges and the corresponding brain signal frequency bands of the four levels of signals by discrete wavelet decomposition.

| Decomposed level | Frequency range (Hz) | Approximate band |
|------------------|----------------------|------------------|
| D1               | 32–64                | Gamma            |
| D2               | 16–32                | Beta             |
| D3               | 8–16                 | Alpha            |
| D4               | 4–8                  | Theta            |

### 2.4 Feature Vector Classification Using Proximal Support Vector Machine (PSVM)

*Classification Method.* In the experiments, we collected data from four difficulty levels (0-, 1-, 2-, 3-back). A popular binary classification technique, support vector machine (SVM), was employed to investigate the data separability at different mental workload levels. SVM techniques have been successfully applied in many classification problems [17–21]. The fundamental problem of SVM is to build an optimal decision boundary to separate two categories of data. Let  $X$  denote a  $n \times k$  dimensional feature vector for a multi-channel data session at certain difficulty level, where  $n$  is the number of signal channels and  $k$  is the number of features of each channel. To classify data between musicians and non-musicians, let  $l$  denote the sample class label and  $l = 1$  denotes musician, and  $l = -1$  means non-musician.

Assume we have  $p$  sessions of level one denoted by  $S_1 = \{(X_1, l_1), (X_2, l_2), \dots, (X_p, l_p)\}$ , and  $q$  sessions of level two denoted by  $S_2 = \{(X_{p+1}, l_{p+1}), (X_{p+2}, l_{p+2}), \dots, (X_{p+q}, l_{p+q})\}$ . Each session is represented by a  $n \times k$  dimensional feature vector. One can find infinitely many hyperplanes in  $R^{n \times k}$  to separate the two data groups.

Standard SVM classifiers, such as Lagrangian Support Vector Machine (LSVM), usually require a large amount of computation time for training. Mangasarian and Wild [22] claims the Proximal SVM (PSVM) algorithm was about 10 to 20 times faster than LSVM. The formulation for the linear PSVM is described as follows:

$$\min_{\omega, \xi, b} \left\{ \frac{1}{2} (\|\omega\|^2 + b^2) + \frac{1}{2} C \xi_i^T \xi_i : D(X^T \omega + b e) = e - \xi_i \right\}, \quad (1)$$

**Table 2.** 19 groups of features are considered. Sub-features are considered in some groups by means of considering various frequency bands and/or different statistics.

| F  | Features (F)                     | Sub-features (f)  |
|----|----------------------------------|---|
| 1  | Basic statistics                 | Average; variance; skewness; kurtosis   |
| 2  | Curve length [8,9]               | Curve length  |
| 3  | Peak count                       | Peak count  |
| 4  | Average nonlinear energy [10,11] | Average nonlinear energy  |
| 5  | Zero crossing                    | Zero crossing   |
| 6  | Spectral edge frequency [12]     | Spectral edge Frequency   |
| 7  | Band power                       | 4 to 8; 8 to 13; 13 to 25; 25 to 40 (Hz)  |
| 8  | Relative band power              | 4 to 8; 8 to 13; 13 to 25; 25 to 40 (Hz)  |
| 9  | Inter-regional asymmetry [13]    | Left vs right channels;<br>Left groups: ([1 3 7], [4 5 6 9], [8 10 11 13]) vs<br>Right groups: ([27 28 30], [24 25 26 29], [18 20 21 23]) |
| 10 | Intra-regional asymmetry [13]    | Left groups: ([1 3 7], [4 5 6 9], [8 10 11 13]) vs<br>Right groups: ([27 28 30], [24 25 26 29], [18 20 21 23])                            |
| 11 | Hurst                            | Hurst   |
| 12 | Hjorth                           | Activity, mobility, complexity  |
| 13 | Barlow                           | Amplitude, frequency, SPI   |
| 14 | Wavelet entropy [14-16]          | Wavelet entropy, power ratio of theta, alpha+low beta,<br>Beta and low gamma, alpha to beta ratio   |
| 15 | Brain rate                       | Brain rate  |
| 16 | Wackermann                       | Sigma, Phi, Omega   |
| 18 | Wavelet statistics               | Max, mean, stdev of bands 4 to 8, 8 to 16, 16 to 32, 32 to 64 (Hz)  |
| 19 | Range to variance ratio          | Range to variance ratio   |
| 20 | Network correlation              | Network correlation   |
| 21 | Approximate entropy              | Approximate entropy   |

where the traditional SVM inequality constraint is replaced by an equality constraint. This modification changes the nature of the support hyperplanes ( $\omega^T X + b = \pm 1$ ). Instead of bounding planes, the hyperplanes of PSVM can be thought of as ‘proximal’ planes, around which the points of each class are clustered and which are pushed as far apart as possible by the term ( $\|\omega\|^2 + b^2$ ) in the above objective function. It has been shown that PSVM has comparable classification performance to that of standard SVM classifiers, but can be an order of magnitude faster [22]. Therefore, we employed PSVM in this study.

*Training and Evaluation.* A classification problem generally follows a two-step procedure which consists of training and testing phases. During the training phase, a classifier is trained to achieve the optimal separation for the training data set. Then in the testing phase, the trained classifier is used to classify new samples with unknown class information. The N-fold cross-validation is an attractive method of model evaluation when the sample size is small. It is capable of providing almost unbiased estimate of the generalization ability of a classifier. For the 29 subjects, the total number of data samples (trials) for session A and B are 192 and 386 respectively. We designed a 2-fold cross-validation method to train and evaluate the SVM classifier [23].

To explore the differences of the responses of musicians and non-musicians under various events, we separate the data into five and three epochs for session

**Table 3.** Based on event labels, 12 and 21 conditions are defined for Group A and Group B respectively. All samples, Pictures only and Words only are considered. In these three subsets, we further split them into cases of Hit, Miss and Correct Rejection. In Group B, among those cases of Hit and Correct Rejection, we further split those subset by the responses, i.e. Low Confidence, High Confidence and New.

| Group A |                                 | Group B |   |
|---------|---------------------------------|---------|---|
| CD      | event                           | CD      | event                                   |
| 1       | All samples                     | 13      | All samples                             |
| 2       | Pictures                        | 14      | Pictures                                |
| 3       | Words                           | 15      | Words                                   |
| 4       | All samples & Hit               | 16      | All samples & Hit                       |
| 5       | All samples & Miss              | 17      | All samples & Miss                      |
| 6       | All samples & Correct Rejection | 18      | All samples & Correct Rejection         |
| 7       | Pictures & Hit                  | 19      | Pictures & Hit                          |
| 8       | Pictures & Miss                 | 20      | Pictures & Miss                         |
| 9       | Pictures & Correct Rejection    | 21      | Pictures & Correct Rejection            |
| 10      | Words & Hit                     | 22      | Words & Hit                             |
| 11      | Words & Miss                    | 23      | Words & Miss                            |
| 12      | Words & Correct Rejection       | 24      | Words & Correct Rejection               |
|         |                                 | 25      | All samples & Low Confidence & Correct  |
|         |                                 | 26      | All samples & High Confidence & Correct |
|         |                                 | 27      | All samples & New & Correct             |
|         |                                 | 28      | Pictures & Low Confidence & Correct     |
|         |                                 | 29      | Pictures & High Confidence & Correct    |
|         |                                 | 30      | Pictures & New & Correct                |
|         |                                 | 31      | Words & Low Confidence & Correct        |
|         |                                 | 32      | Words & High Confidence & Correct       |
|         |                                 | 33      | Words & New & Correct                   |

A and B respectively based on the test phases as shown in Fig. 1. In addition, A3 is further separated into five pieces with one second for each piece in order to study various parts of A3. We also study the first  $l \in 0.4, 0.6, 0.8, 1, 1.5$  s of B2. These subsegments are denoted as A21, A22, ..., A25 and B21, B22, ..., B25. Based on the event markers of the EEG data, we define 12 conditions for session A and 21 conditions for session B. The following table lists all of the conditions.

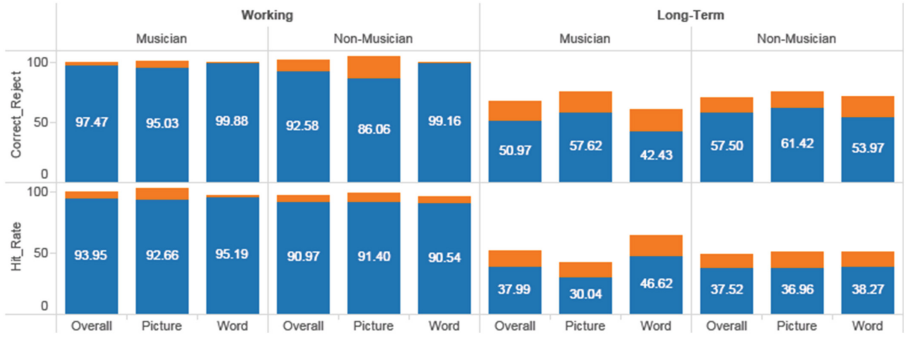
In testing, for each comparison group, we divided the corresponding data samples into 5 non-overlapping subsets. Each time we picked one subset out and trained the PSVM classifier by the data samples of another set. The samples of the left-out subset were considered as unknown samples to test the performance of the trained classifiers. Repeating this procedure again for another set, the averaged prediction accuracy over the 5-fold runs was used to indicate the degree of separability of the EEG signals of musicians and non-musicians.

To achieve reliable feature selection, we employed an advanced feature selection technique, called minimum redundancy maximum relevance (mRMR) [24], which allows us to select a subset of superior features at a low computational cost in a high dimensional space.

The basic idea of mRMR is to select the most relevant features with respect to class labels while minimizing redundancy amongst the selected features. The mRMR algorithm uses mutual information as a distance measure to compute feature-to-feature and feature-to-class-label non-linear similarities.

### 3 Experimental Results

Before going into the classification results, Fig. 2 shows percentage of hit rate, correct rejection rate and the corresponding standard deviation. We noted that musicians had higher hit rate on picture than non-musicians. Also, musicians performed better in working memory task but they performed worse in long term memory task than non-musicians.



**Fig. 2.** This is a boxplot of AUC of all models using only one sub-feature at a time among various epochs aggregated by condition. For each box, there are 68 sub-features, 8 to 10 epochs and so there are roughly 700 results. Obviously, working memory (condition 1 to 12) have better classification results than long term memory.

Table 4 is a summary of classification accuracy on various conditions. On response period (B2) of long-term memory data as well as maintain period (A3) and response period (A4) of working memory, we obtained high classification accuracy. We also observed that miss events obtained high classification accuracy in general.

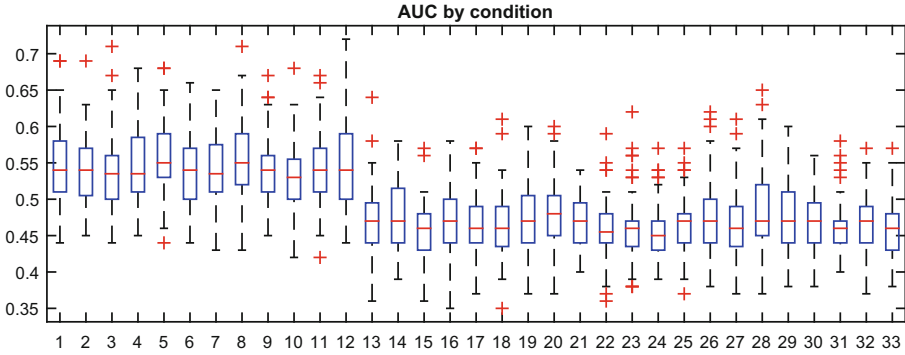
**Table 4.** Summary of classification accuracy on various conditions and epochs in working and long-term memory tasks.

| Conditions          | Study period (working memory) |       |          |          |           | Test period (long-term memory) |          |           |
|---------------------|-------------------------------|-------|----------|----------|-----------|--------------------------------|----------|-----------|
|                     | A1                            | A2    | A3       | A4       | A5        | B1                             | B2       | B3        |
|                     | Pre-stim                      | Stim1 | Maintain | Response | Post-stim | Pre-stim                       | Response | Post-stim |
| All samples         | 83.78                         | 70.27 | 83.78    | 83.78    | 72.97     | 66.67                          | 75       | 69.44     |
| Pictures            | 75.68                         | 70.27 | 72.97    | 72.97    | 67.57     | 78.78                          | 75       | 72.22     |
| Words               | 81.08                         | 72.97 | 86.49    | 78.38    | 64.86     | 88.89                          | 69.44    | 62.86     |
| All hit             | 64.86                         | 72.97 | 81.08    | 78.37    | 72.97     | 87.5                           | 91.67    | 64.29     |
| All - miss          | 64.86                         | 70.27 | 83.78    | 83.78    | 72.97     | 82.61                          | 83.33    | 90.63     |
| All corr. rej       | 89.19                         | 67.57 | 78.35    | 83.78    | 67.57     | 52.17                          | 88.89    | 80.65     |
| Picture hit         | 78.38                         | 78.38 | 67.57    | 70.27    | 70.27     | 63.64                          | 75       | 73.53     |
| Picture - miss      | 78.38                         | 83.78 | 72.97    | 72.97    | 78.38     | 73.68                          | 75       | 72.41     |
| Picture - corr. rej | 62.16                         | 70.27 | 72.97    | 64.86    | 59.46     | 79.95                          | 72.22    | 80.65     |
| Word hit            | 59.46                         | 64.86 | 75.68    | 67.57    | 67.57     | 55                             | 86.11    | 68.75     |
| Word - miss         | 59.46                         | 78.38 | 64.86    | 70.27    | 78.38     | 73.68                          | 77.78    | 80        |
| Word corr. rej      | 72.97                         | 70.27 | 83.78    | 67.57    | 64.86     | 84.21                          | 69.44    | 75.86     |

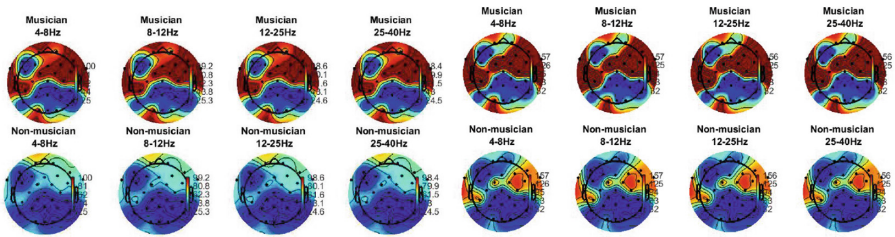
Figure 3 shows that conditions 1 to 12 (Group A) have higher area under the curve (AUC). It is obvious that EEG of musicians and non-musicians have the most difference during short term memory task.

Figure 4 are the topographies of band power of musicians (left) and non-musicians (right). We note that musicians have larger range of bandpower values than non-musicians. Also, musicians tend to be active on multiple locations while

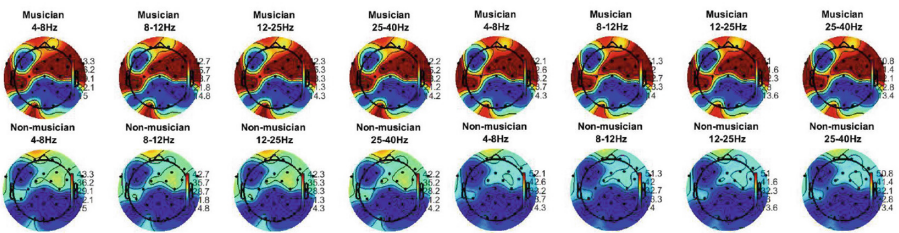




**Fig. 3.** This is a boxplot of AUC of all models using only one sub-feature at a time among various epochs aggregated by condition. For each box, there are 68 sub-features, 8 to 10 epochs and so there are roughly 700 results. Obviously, working memory (condition 1 to 12) have better classification results than long term memory.



(a) Maintain Period in Working Memory for Miss conditions (b) Maintain Period in Working Memory for Hit conditions



(c) Response Period in Long Term memory for Miss conditions (d) Response Period in Long-Term memory for All Hit

**Fig. 4.** Comparison for averaged 4 EEG band power between musicians and non-musicians in response and maintain period. In all conditions, musicians demonstrate higher level of activity in frontal area. Their values are about 50% higher than non-musicians. Musicians are also more active in several more areas (right area, central area and left middle area) while non-musicians mainly only use their frontal area. This pattern is more obvious in long term memory.

non-musicians tend to be active only on the front lobe. In this study, we also investigated the effect of ICA artifact removal and found that the effect was significant. It improves the classification accuracy by 10% in general.

## 4 Conclusion and Future Work

In this study, we made a comprehensive EEG data mining study to investigate and compare cognitive memory processing for musicians and non-musicians. We presented a computational EEG pattern analysis and classification framework, which integrated the most recent advances in automated spatiotemporal artifact removal, a broad selection of most popular EEG feature extraction techniques, an information-theory-based feature selection, and a PSVM classification model. The experimental results show that the EEG patterns of the active memory encoding process at the maintain period indeed demonstrated significant differences between musicians and non-musicians. Our study found that musicians overall demonstrated better and more accurate memory performance in both short-term and long-term memory tasks. In particular, the EEG brainwave differences of musicians were more significant on the short-term memory tasks compared to the non-musician group. From the four common EEG band power study, we noted that the musicians were significant more active in frontal areas in alpha, beta, and low gamma bands than non-musicians. This may indicate that the long-time music training can sharpen brain pathways in memory processing with a more active brain activity during memory tasks. More analysis on EEG spatiotemporal patterns and memory brain network will be investigated in future works. The integrated computational framework developed in this study also provides a powerful tool to perform EEG signal processing and pattern analysis, and can be useful in many other applications that involve pattern recognition or abnormality detection in multivariate EEG signals.

## References

1. Lin, Y.-P., Wang, C.-H., Jung, T.-P., Wu, T.-L., Jeng, S.-K., Duann, J.-R., Chen, J.-H.: EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* **57**(7), 1798–1806 (2010)
2. Murugappan, M., Nagarajan, R., Yaacob, S.: Comparison of different wavelet features from EEG signals for classifying human emotions. In: *IEEE Symposium on Industrial Electronics and Applications ISIEA 2009*, vol. 2, pp. 836–841. IEEE (2009)
3. Lin, Y.-P., Wang, C.-H., Wu, T.-L., Jeng, S.-K., Chen, J.-H.: Support vector machine for EEG signal classification during listening to emotional music. In: *2008 IEEE 10th Workshop on Multimedia Signal Processing*, pp. 127–130. IEEE (2008)
4. Croft, R., Barry, R.: Removal of ocular artifact from the EEG: a review. *Clin. Neurophysiol.* **30**, 5–19 (2000)
5. Mognon, A., Jovicich, J., Bruzzone, L., Buiatti, M.: Adjust: an automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* **48**(2), 229–240 (2011)
6. Wang, S., Lin, C., Wu, C., Chaovalitwongse, W.: Early detection of numerical typing errors using data mining techniques. *IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum.* **41**(6), 1199–1212 (2011)

7. Grimes, D., Tan, D., Hudson, S., Shenoy, P., Rao, R.: Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, vol. 08, pp. 835–844 (2008)
8. Olsen, D., Lesser, R., Harris, J., Webber, R., Cristion, J.: Automatic detection of seizures using electroencephalographic signals, U.S. Patent 5311876
9. Esteller, R., Echauz, J., Cheng, T., Litt, B., Pless, B.: An efficient feature for seizure onset detection. In: Proceedings of the 23rd International Conference of IEEE Engineering Medicine Biology Society, vol. 2, pp. 1707–1710 (2001)
10. Kaiser, J.: On a simple algorithm to calculate the energy of a signal. In: Proceedings of 1990 International Conference of Acoustics, Speech, Signal Processing, vol. 1, pp. 381–384 (1990)
11. Agarwal, R., Gotman, J.: Adaptive segmentation of electroencephalographic data using a nonlinear energy operator. In: Proceedings of 1999 IEEE International Symposium on Circuits and Systems, vol. 4, pp. 199–202 (1999)
12. Addison, N.: The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine, and finance, Taylor and Francis
13. Yuvaraj, R., Murugappan, M., Ibrahim, N.M., Omar, M.I., Sundaraj, K., Mohamad, K., Palaniappan, R., Mesquita, E., Satiyan, M.: On the analysis of EEG power, frequency and asymmetry in Parkinsons disease during emotion processing. *Behav. Brain Func.* **10**(1), 12 (2014)
14. Rosso, O., Martin, M., Figliola, A., Keller, K., Plastino, A.: EEG analysis using wavelet-based information tools. *J. Neurosci. Methods* **153**(2), 163–182 (2006)
15. Subasi, A.: EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst. Appl.* **32**(4), 1084–1093 (2007)
16. Rosso, O., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schourmann, M., Basar, E.: Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *J. Neurosci. Methods* **105**(1), 65–75 (2001)
17. Blankertz, B., Curio, G., Muller, K.: Classifying single trial EEG: towards brain computer interfacing. *Adv. Neural Inf. Process. Syst.* **14**(2), 157–164 (2002)
18. Lal, T., Hinterberger, T., Widman, G., Schroer, M., Hill, J., Rosenstiel, W., Elger, C., Schokopf, B., Birbaumer, N.: Methods towards invasive human brain computer interfaces. In: *Advances in Neural Information Processing Systems*, vol. 17, pp. 737–744. MIT Press (2005)
19. Rakotomamonjy, A., Guigue, V., Mallet, G., Alvarado, V.: Ensemble of SVMs for improving brain computer interface P300 speller performances. In: Duch, W., Kacprzyk, J., Oja, E., Zadrozny, S. (eds.) *ICANN 2005. LNCS*, vol. 3696, pp. 45–50. Springer, Heidelberg (2005). doi:[10.1007/11550822\\_8](https://doi.org/10.1007/11550822_8)
20. Kaper, M., Meinicke, P., Grossekhoefer, U., Lingner, T., Ritter, H.: BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm. *IEEE Trans. Biomed. Eng.* **51**(6), 1073–1076 (2004)
21. Garrett, D., Peterson, D., Anderson, C., Thaut, M.: Comparison of linear, non-linear, and feature selection methods for EEG signal classification. *IEEE Trans. Neural Syst.Rehabil. Eng.* **11**(2), 141–144 (2003)
22. Mangasarian, O., Wild, E.: Proximal support vector machine classifiers. In: *Proceedings of Knowledge Discovery and Data Mining*, pp. 77–86 (2001)
23. Stone, M.: Cross-validatory choice and assessment of statistical predictions. *J. Roy. Stat. Soc.: Ser. B (Statistical Methodological)* **36**(2), 111–147 (1974)
24. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(8), 1226–1238 (2005)