Assessing Neuroplasticity with Convolutional and Recurrent Neural Networks

Irán Román
Stanford University
660 Lomita Dr, Stanford, CA 94305
iran@stanford.edu

Abstract

When humans receive musical training, the neural substrates associated with production and processing of music undergo functional changes. This study investigates whether a Convolutional Neural Network (CNN) can identify changes in the neural dynamics of an adult that received musical training. Two different CCN architectures were trained on Magnetoencephalography (MEG) time-series from seven adults before they received musical lessons. These subjects listened a steady metronome beat while their brain activity was recorded. The classification task for the CNN was to identify data belonging to a specific subject. After training, each CNN was presented with MEG data from the same subjects listening the same metronome beat, but recorded after they took musical lesson three times per week for five weeks. CNNs were trained with different data transformations, which included: the entire dataset, spatial clustering, and PCA decomposition. Data was also fit to the hidden modules of a Clockwork Recurrent Neural Network.

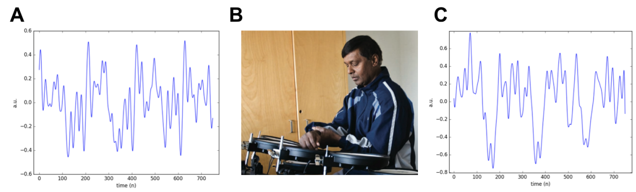

Figure 1. (A) The first principal component of MEG data recorded from a subject listening an isochronous metronome beat before he underwent musical training. (B) A subject undergoing multi-modal musical training. (C) The first principal component for MEG data from the same subject in (A) listening the same metronome beat after experiencing five weeks of intensive musical training. Notice, for example, the greater relative amplitude and more uniform envelope of the 1st principal component post-training, compared to before training.

1. Introduction

When humans learn or recover an ability throughout life, the brain undergoes functional and anatomical changes. These transformations in the brain are known as "Neuroplastic" changes, and occur in order to allow the brain to optimally carry out novel tasks [12]. Although there is strong evidence suggesting that the brain undergoes changes driven by learning of a specific skill, the potential interpretation of neuroplasticity in therapeutical and education settings remains a challenge due to the lack of a standarized and reliable method for the assessment of neuroplasticity at the individual and group levels [9].

Musical training is a specific kind of skill learning that engages the individual multimodaly, and leads to functional and anatomical changes in both cortical and subcortical brain areas. A 2012 study [3] pointed out how brain oscillatory activity in the beta band predicts metronome beats and changes after subjects receive musical training.

Several human non-invasive methods to record brain activity exist today. They have different advantages and disadvantages, with some of them having better spatial resolution than others, such as fMRI vs Electroencephalography (EEG), and others with better temporal resolution, like Magnetoencephalography (MEG) and EEG. Aided with a source estimation and reconstruction algorithm, MEG presents itself as the option with the highest temporal and spatial resolution available. Thus, this study used MEG time-series data from seven subjects listening an isochronous metronome beat before and after they experienced intensive musical training for five weeks. The effect of musical training on one subjects can be observed in figure 1.

The goal of this investigation was to build a Convolutional Neural Network (CNN) that takes time-series data from the different voxels in the head and classifies it as belonging to a specific subject. The CNN would be trained on data before the subjects underwent musical training. The performance of the trained CNN on the data from the same subjects after they received intensive musical lessons would be used to assess changes in the brain of each subject after musical training. This is a novel application of CNNs that could lead to novel and reliable methods for the assessment of neuroplasticity in terms of the temporal dynamics across different brain areas.

2. Related Work

As observed in [11], human brain recordings collected longitudinally for the assessment of training-induced neuroplasticity are currently assessed by physicians and neuroscientists using standard statistical tools applied to manually-selected features. Thus, current methodologies return highly variant results, ofen open to a number of interpretations. Novel neural networks could be used to fine-tune the analysis of brain data with the goal of objective assessment of neuroplasticity.

Previous studies have used CNN as classifiers to identify patterns of brain activity in time-series data and tell apart activity generated by different individuals [8]. A pioneering study developed CNNs to identify a brain potential commonly used in Brain-Computer-Interfaces from electric brain activity recorded with EEG. Their best CNN was able to recognize the brain potential at a rate of 95.5 percent [10]. This study used hevily pre-processed brain data where noisy channels were rejected and unwanted spectral components were filtered out before passing the data into the CNN. A more exploratory study [14] asked whether CNNs could be robust enough to abstract features in noisier data. These CNNs took time-series brain data that was not manually cleaned before entering the CNNs. Performance was less successful (best test accuracy was 50 percent) towards the goal of classifying brain activity as triggered by a specific rhythmic stimulus.

Temporal dynamics across time scales reflect different important component of time-series data from the brain. For this purpose, Long-Short-Term-Memory (LSTM) and Recurrent Neural Networks have been used to capture the features of brain dynamics with a higher temporal resolution [2]. The Clockwork Recurrent Neural Network (CRNN) [6] is a recently developed algorithm, able to capture the structure of time-series sequences with finer temporal detail, compared to other Recurrent Neural Networks (RNN) or LSTMs, due to the activation of its hidden units at different time steps. More CRNN theory is described in the methods section 3.2.2.

Even though still in their early stages of development as tools for the analysis of time-series brain data, CNNs and RNNs are attractive models that can be used to assess neuroplastic changes in the brain driven by musical engagement, and as observed in the dynamics of brain activity in individuals before and after musical training.

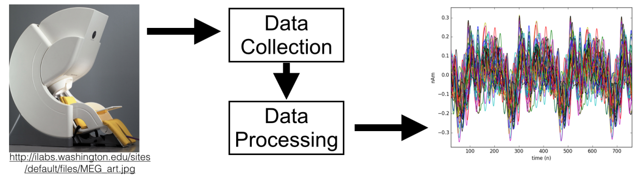

Figure 2. An MEG machine (left). A coarse illustration of the pipeline for data collection and processing (center). Example data as an overlay of the time-series for all voxels in a subject (right).

3. Methods

3.1. Data Collection and Preprocessing

Seven healthy male subjects with normal hearing were recruited to participate for data collection. They were presented with a isochronus metronome beat appearing every 800 ms. Their brain activity was recorded with an MEG machine (see figure 2), which captures magnetic fields that are produced by the electric currents of the brain. After an initial MEG recording, these subjects received, for a period of five weeks, musical training through which they learned to play piano, sing, play percussion instruments, and music theory. This training protocol was motivated by the fact that music activities of this kind lead to engagement of several functional networks in the brain including auditory, motor, language, visual, memory, and attention. Thus, this training protocol will lead to the recruitment and strengthening of neural pathways between the brain modules required for efficient musical action production. After this period of musical training, each subject's brain activity to the same metronome beat was recorded using MEG.

Since performance of CNNs with preprocessed data has been higher in previous studies, data here was also preprocessed. MEG data for each subject were transformed from 150 MEG coils collecting magnetic fields at the scalp, to multiple voxels inside the entire head with discrete sources of time-series acticity that collectively generate the magnetic fields observed at the scalp. This is known as the beamforming algorithm [13]. After beamforming, each subject had over 1000 (8mm cubes) spatial voxels of MEG activity, each containing time-series activity at a sampling rate of 2,083 samples per second. MEG data were downsampled to 260 samples per second, and from this downsampled data, the beta-band power for each voxel was calculated using a band-pass filter to isolate the beta-band (15Hz to 30Hz), followed by a Hilbert Transform.

Given the different head shapes across subjects, only the voxels that overlaped across subjects were kept as input to the classification algorithms. Thus, the overlapping space across subject was a sphere of voxels with the shape observed in figure 3.

Additionally, data was normalized to have zero mean

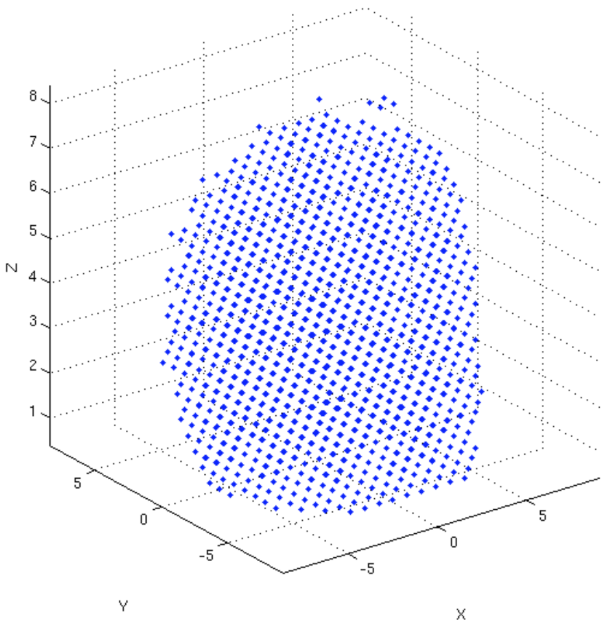

Figure 3. Spherical shape consisting of voxel overlap across seven subjects for both MEG recordings. Voxel location is determined by the head-based coordinates: the X axis is the anterior (positive) and posterior (negative) sides of the head, the Y axis is the right (negative) and left (positive) hemispheres, and Z is the superior (positive) and inferior (negative) directions of the head.

and unit variance across subjects. Networks were trained to identify subjects only on the pre-musical training data set. The post-musical training data set was used to test subject identification performance by the same neural network, but as data that reflects the electrophisiolgical changes induced by musical training.

3.2. Neural Networks

This investigation studied the MEG data using two types of Neural Networks, convolutional and clockwork recurrent. Here is a description of the theory and software implementation for each.

3.2.1 Convolutional Neural Networks

CNNs were built using theano [1] and lasagne[15]. These libraries combined allow for the training of neural networks with one-dimensional time-series data.

The first round of experiments with CNNs focused on the classification of subjects based on time-series activity from head voxels. The architecture was selected as the one that achieved a classification accruracy of at least 85 percent on the test set after training of the network. The selected architecture, represented in figure **??**, consisted of three convolutional layers, each followed by a max-pooling

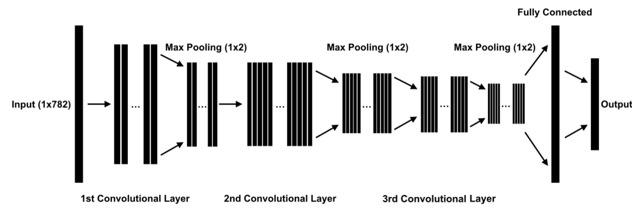

Figure 4. CNN architecture for classification of subjects based on time-series activity from head voxels. The network contained a sequence of three convolutional layers, each followed by a max-pooling layer. Afther convolutions and poolings, the network features a fully-connected layer with a ReLU non-linearity and an output layer with softmax activation.

layer, leading to a fully-connected layer, which was connected to the output layer. All convolutional layers used a stride equal of 1 and no zero padding. The first convolutional layer carried out convolutions on the input data (size 1x782) using 100 filters, each of 1x100 size. All operations kept the one-dimensional quality of the input data. A max pooling operation of size 1x2 on the output of the first convolutional layer was carried out. After max pooling, a second convolutional layer applied 50 filters of size 1x50 each to the output of the first max pooling. Another max pooling operator of size 1x2 was applied to the output of the second convolutional layer. After max pooling, a third convolutional layer applied 25 filters of size 1x25 to the output of the second max pooling layer. Another max pooling operator of size 1x2 was applied to the output of the third convolutional layer. This chain of three convolution-max pooling operations was followed by a fully connected layer with 512 units activated by a ReLU non-linearity computing the maximum between zero and the input to a unit in the fully connected layer $f(x) = max(0, x)$, then leading to the output layer consisting of a softmax function

$$f_j(z) = \frac{ez_j}{\sum_{i=1}^{k} e^{z_i}}$$

returning the probability for each of the multiple classes (k different ones), which for the purpose of this study will be subjet number to whom the input data belonged. The loss function for softmax classification, and thus the output of the network, is given by the cross entropy loss, which has the form

$$L_i = -f_{y_i} + log(\sum_{j=1}^{k} e^{f_j})$$

where the function f is the softmax function.

The second round of experiments with CNNs pursued the classification of subjects based on the first ten principal components obtained through PCA analysis of the time-series data for all voxels. The equation for PCA that finds the first principal component is found through the maximization of

$$\frac{1}{m}\sum_{i=1}^{m}(x^{(i)^T}u)^2 = u^T\left(\frac{1}{m}\sum_{i=1}^{m}x^{(i)}x^{(i)^T}\right)u$$

where u is the unit vector representing the direction on which the data points in $x^{(i)}$ can be projected. The first principal component found by PCA depicts the axis on which the variance of the data is retained and maximized. We can find k principal components by maximizing the first k u's to project the data onto a k-dimensional subspace. All the principal compoents for the time-series data across different voxels can thus be found by the expression

$$\mathbf{W} = \mathbf{XU}$$

where U is a p-by-p matrix whose columns are the eigenvectors of X^TX, and W is a matrix containing the PCA analysis of the entire dataset.

The architecture for this second round of experiments is shown in figure 5, and was also selected as the one that achieved the best classification accruacy, which in this case was only of 76 percent on the test set. This architecture is similar to the one in figure 4, with the main difference of having an extra convolutional layer and using different number of filters and filter sizes in each convolutional layer. All convolutional layers used a stride equal of 1 and no zero padding. The first convolutional layer carried out convolutions on the input data (size 1x782) using 500 filters, each of 1x50 size. The second convolutional layer applied 200 filters of size 1x50 each to the output of the first max pooling. The third convolutional layer applied 100 filters of size 1x50 to the output of the second max pooling layer. The fourth convolutional layer applied 50 filters of size 1x25 to the output of the third max pooling layer. Another max pooling operator of size 1x2 was applied to the output of the fourth convolutional layer. This chain of four convolution-max pooling operations was followed by a fully connected layer with 512 units activated by a ReLU non-linearity, and an output layer with a softmax function.

3.2.2 Clockwork Recurrent Neural Network

Finally, this study also attempted the implementation of an RNN called the Clockwork Recurrent Neural Network (CRNN). This type of recurrent network has the advantage of mantaining a high-temporal-resolution memory in its hidden layers after training. The main advantage of this network is the fact that it overcomes the problem of the vanishing gradient found in other RNNs by partitioning the neurons in its hidden layers as different "sub-clocks" that are able to capture the input to the network at different time steps. The output of this neural network at time-step t is described in [6] by the equations

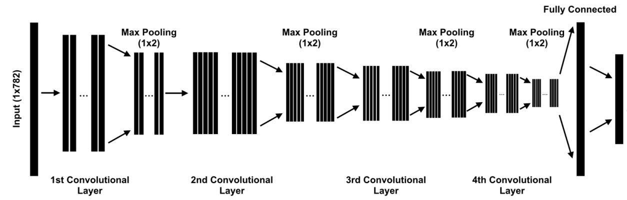

Figure 5. CNN architecture for classification of subjects based on the first ten principal pomponents as obtained by PCA decomposition of the time-series activity from head voxels. The network contained a sequence of four convolutional layers, each followed by a max-pooling layer. Afther convolutions and poolings, the network features a fully-connected layer with a ReLU non-linearity and an output layer with softmax activation.

$$y_H^{(t)} = f_H(W_H \cdot y^{(t-1)} + W_L \cdot x^{(t)})$$

$$y_O^{(t)} = f_O(W_O \cdot y_H^{(t-1)})$$

where the subindices I, H, and O, indicate elements in the input, hidden, and output layers. $x^(t)$ is the input of the network, and f describes activation functions for the different steps of the network.

The Implementation of this network used Torch [16] and was developed by modifying a version found at [4]. The architecture of this network featured an input, connected to seven hidden layers. All hidden layers had seven "subclocks", which are submodules in each hidden layer featuring different cycle lengths. Within each module, all the clocks in the hidden layer are fully connected, and depending on the clock period, the clocks of a faster module i are connected to the clocks in a slower module j if the clock period $T_i < T_j$.

Given the time-series nature of the data for this study, and the fact that performance of the Clockwork Recurrent Neural network is superior to the performance of standard RNNs and LSTMS, this network could optimally capture temporal dynamics and functional changes in the brain after subjects received intense musical lessons.

4. Description of Experiments

Now, a description of the four main studies carried out in this investigation is presented.

4.1. Classification of Subjects with the Full Dataset

The network architecture described in figure 4 was trained to classify subjects using all the data recorded before the subjects received musical lessons. The data was divided into training, validation, and test sets by randomly separating 100 voxels across subjects for validation, 100 for testing, and keeping the remaining ones for training. The

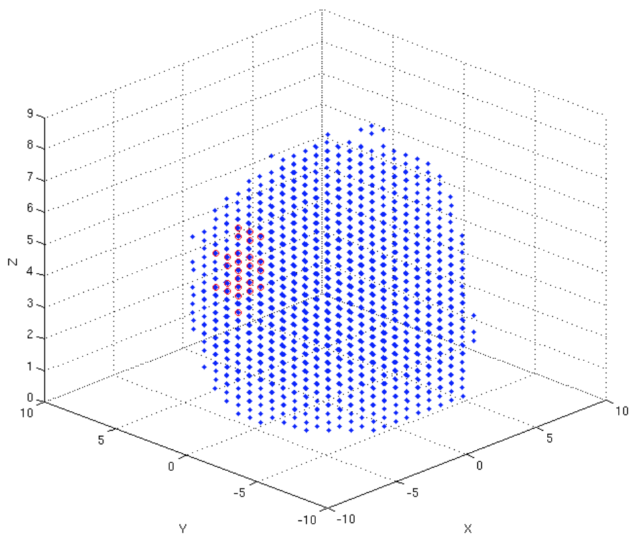

Figure 6. Example of a cluster of 25 nearest-neighbor voxels (red) selected from the entire set of voxels from the 3D head shape (blue).

learning rate used to carry out Stochastic Gradient Descent was 5e-3, weight initialization was done using the method described in [5], and Nesterov momentum was used with a constant μ of 0.9. Performance of the network was assessed by dividing the training and validation sets into batches and repeating training through 10 epochs.

After training the network and achieving a test accuracy over 90 percent, the same network was presented with 100 voxels randomly selected across subjects from the MEG recording after all subjects received intense musical lessons for five weeks.

4.2. Classification of Subjects with Voxel Clusters

The same network architecture used in the previous experiment was presented with different nearest-neighbor clusters of 25 voxels grouped and selected from all over the head 3D space as seen in figure 6. Data to train the network was again taken from the MEG recording before the subjects received musical lessons. Out of the total 25 voxels, 15 voxels in each cluster were randomly selected to be the training set, 5 of the remaining 10 voxels were randomly selected to be the validation set, and the remaining five served as the test set. All training parameters (learning rate, weight initialization, number of epochs, and momentum) remained the same.

After training the network and achieving test accuracies over 85 percent across voxel clusters, and an average test accruacy greater than 90 percent accounting for all clusters, the same network was presented with 12 voxels randomly selected from each cluster from the MEG recording after all subjects received intense musical lessons for five weeks.

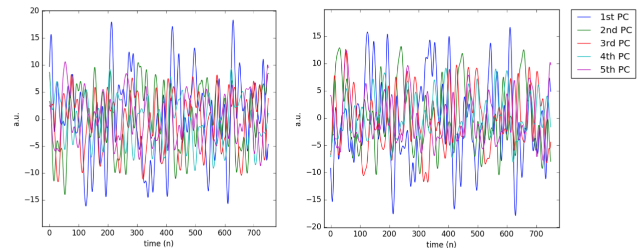

Figure 7. The first five principal components for a single subject after carrying out PCA. Left: before receiving musical trainig. Right: after receiving musical training.

4.3. Classification of Subjects with the First 10 Principal Components of PCA Analysis

Studies on neuroplasticity often assess changes in the brain by looking at transformations of multidimensional data to assess changes in the functional and temporal dynamics that are the outcome of a specific training. Here, the network described in figure 5 was presented with the first 10 principal components obtained after carrying out PCA on the entire data set for each subject (see figure 7). The PCA analysis obtained on the subject data before they underwent musical lessons was used to train the network. The classification task was to determine whether data belonged to a specific subject. Two thirds of the PCA data available for the first ten principal components were randomly selected to be used for training the network, one sixth of the remaining data was randomly selected to serve as the validation set, and the remaining sixth of the data served as the test set. The learning rate used to carry out Stochastic Gradient descent was 5e-4, weight initialization was done using the He's method, and Nesterov momentum was used with a constant μ of 0.5. Performance of this network was also assessed by dividing the training and validation sets into batches and repeating training through 15 epochs.

After training the network and achieving the best test accuracy possible, which in this case was 76 percent, the same network was presented with the first ten principal components obtained after carrying out PCA on the data obtained from the same subjects after they took musical lessons for five weeks.

4.4. Fitting the Data to a CRNN

The first principal component obtained through PCA on a randomly selected subject before musical lessons was the input to a CRNN implementation modified from [4]. The number of hidden units was 7, the learning rate was 3e-5 with nesterov momentum of 0.99, and weight initialization drawn from a uniform distribution. The activation function was tanh throughout the network.

This is work in progress, and only the current best training case is reported in the next section.

	Data Set	Accuracy
Experiment 1	Validation	92.14%
	Test	91.14%
	Post-Training	29.86%
Experiment 3	Validation	56.67%
	Test	76.67%
	Post-Training	21.43%

Figure 8. Summary of results for experiments 4.1 and 4.3.

5. Discussion of Results

Now, the main results for these four main studies are presented.

5.1. Classification of Subjects with the Full Dataset

The network achieved a validation accuracy over 92 percent classifying data as belonging to a specific subject, and was able to achieve a test set accuracy over 91 percent. However, when presented with the data from the same subjects after they received musical lessons, the network only achieved an accuracy of 30 percent. This clearly suggests that the brain of these subjects underwent functional changes during the period that they received musical training. Results for this experiment are summarized in figure 8.

Since this experiment used data from the entire head of the subjects, and the functional changes are likely not homogeneous throughout the head, the procedure of the next experiment was motivated.

5.2. Classification of Subjects with Voxel Clusters

The results for this experiment are sumarized in figure 9A. Figure 9B features example plots of the training loss for 10 randomly selected clusters. All voxel clusters led to accuracies on the test of at least 85 percent. Accuracies on the data set recorded after subjects received musical training were more diverse, which motivates interest for spatial visualization of these results to assess which brain regions could be associated with different accruacies. Subject classification based on voxel clusters allows for the separation of regions given their associated classification accruacy on the data set recorded from subjects after receiving musical lessons. A visualization of these results is presented in figure 10.

Observing these clusters and their associated accuracies on the post-musical training data set, we can estimate which brain regions changed more their functional activity than others. Regions with a higher accuracy on the post-musical training data set could suggest brain areas that changed the least, which lower accuracies could indicate brain areas that changed the most.

A

Cluster Data Set	Average Accuracy	Best Accuracy
Validation	~95%	100%
Test	~90%	100%
Post-Training	~20%	43.25%

B

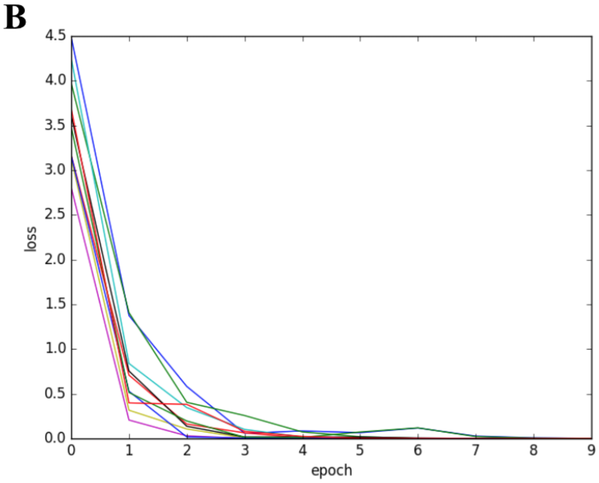

Figure 9. (A) Summary of training, test, and post-training accuracies obtained in experiment 4.2 across voxel clusters. (B) training loss plots for 10 randomly selected voxels.

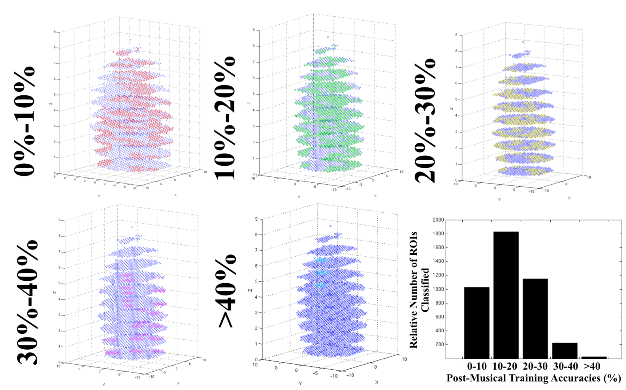

Figure 10. Visualization of the clusters associated with the different accuracies obtained on the post-musical training data set. Only one cluster returned an accruacy on the post-musical training set greater than 40 percent. The histogram in the right bottom corner illustrates the relative number of ROIs (Regions of Interest, i.e. Voxels) that fell under different percent accuracy ranges.

5.3. Classification of Subjects with 10 PCAs

Classification of subjects using the first ten principal components obtained after PCA decomposition was a harder task. The network was only able to reach training accuracies little over 75 percent. In the best training case, the validation accuracy was 56 percent and the accruacy on the test set was 76 percent. The results are summarized in figure 8. Thus, results indicate that the PCAs may be

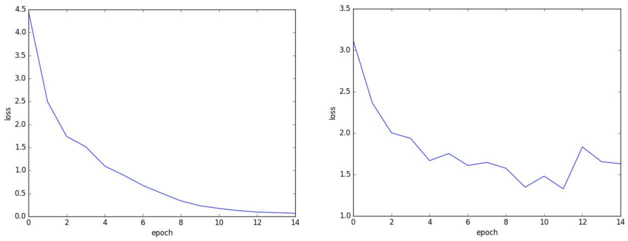

Figure 11. Plots for the training (left) and validation (right) loss for experiment 4.3.

more homogeneous across subjects and harder to extract features from with the goal of subject classification. Figure 11 presents an example of the training and validation loss for this experiment.

Additionally, when presented with the first 10 principal components from the data set collected after subjects received musical training, the network only returned an accuracy little over 21 percent. This drop in performance suggests that the first 10 principal components do contain information about the functional changes in the brain that subjects underwent during the period of musical lessons. However, a finer search for the features that are more consistent or more varying in the principal components across subjects is necessary to improve this experimental approach.

5.4. Fitting the Data to a CRNN

This attempt to fit the first principal component from PCA decomposition of data from a single subject to the hidden modules of a Clockwork RNN has yielded inconclusive but promising results. As seen in figure 12, this network implementation can learn the time-series sequence for the training example that it is fed in. However, this implementation has been very slow and major changes to the algorithm will have to be carried out in order for this network to realistically take more than a single training example. A first step toward the improvement of this algorithm could be as simple as downsampling the data that goes into the network.

Results from [6] seem very promising and motivate the use of CRNNs for the assessment of the data in this investigation, as classification of sequences in their CRNN version outperforms performance of most RNNs and LSTMS.

6. Conclusions

This study has proposed CNN and RNNs as tools for the assessment structural and functional neuroplasticity as observed in MEG data recorded before and after subjects experienced musical training. This study's implementation demonstrates that the architectures and methods proposed can provide insight about the changes that the brain experiences through skill learning. Performance of the methods here presented must be compared to the performance of other classification methods, such as KNN, SVMs, and Softmax before going forward. To expand on experiments described in 4.1 and 4.2, single subject data from two neighboring clusters of voxels could be classified as belonging or not to a single cluster in order to assess similarity between regions of the brain that are near to each other. To expand the results for the experiment discussed in 4.3, not the full 10 first principal components can be used for classification, but perhaps only the ones that are found to be most different between subjects. A full implementation of the Clockwork Recurrent Neural Network experiment is also necessary in order to be able to compare its performance with that of other methods.

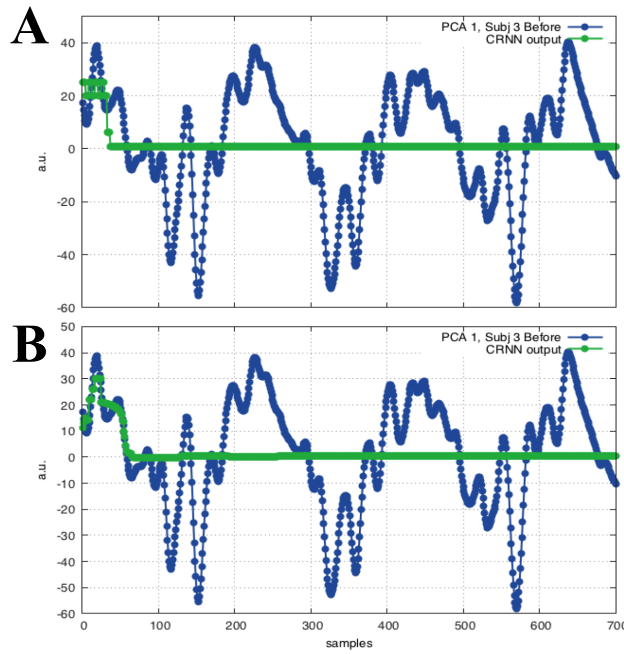

Figure 12. (A) Early training of a Clockwork Recurrent Neural Network on PC 1 from a single subject. (B) Mid training on PC 1 from a single subject with the same CRNN.

Finally, other types of neural networks, such as Gradient Frequency Neural Networks [7], may be better able to capture the changes of temporal dynamic properties of the brain, as captured by MEG. The proposed assessment of neuroplasticity by means of neural networks will expand the methods here described by including Gradient Frequency Neural Networks.

References

[1] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley, and Y. Bengio. Theano: new features and speed improvements. *arXiv preprint arXiv:1211.5590*, 2012.

[2] P. R. Davidson, R. D. Jones, and M. T. Peiris. Eeg-based lapse detection with high temporal resolution. *Biomedical Engineering, IEEE Transactions on*, 54(5):832–839, 2007.

[3] T. Fujioka, J. E. Ween, S. Jamali, D. T. Stuss, and B. Ross. Changes in neuromagnetic beta-band oscillation after music-supported stroke rehabilitation. *Annals of the New York Academy of Sciences*, 1252(1):294–304, 2012.

[4] S. Hansen. Clockwork rnn. `https://github.com/zergylord/ClockworkRNN`, 2015.

[5] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1026–1034, 2015.

[6] J. Koutnik, K. Greff, F. Gomez, and J. Schmidhuber. A clockwork rnn. *arXiv preprint arXiv:1402.3511*, 2014.

[7] E. W. Large, F. V. Almonte, and M. J. Velasco. A canonical model for gradient frequency neural networks. *Physica D: Nonlinear Phenomena*, 239(12):905–911, 2010.

[8] L. Ma, J. W. Minett, T. Blu, and W. S.-Y. Wang. Resting state eeg-based biometrics for individual identification using convolutional neural networks. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 2848–2851. IEEE, 2015.

[9] M. N. McDonnell, S. Koblar, N. S. Ward, J. C. Rothwell, B. Hordacre, and M. C. Ridding. An investigation of cortical neuroplasticity following stroke in adults: is there evidence for a critical window for rehabilitation? *BMC neurology*, 15(1):109, 2015.

[10] F. Nijboer, E. Sellers, J. Mellinger, M. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. Krusienski, T. Vaughan, et al. A p300-based brain–computer interface for people with amyotrophic lateral sclerosis. *Clinical neurophysiology*, 119(8):1909–1916, 2008.

[11] R. Rajapakse and A. Rubasinghe. An experimentation framework towards longitudinal analysis of eeg/erp data to assess brain plasticity. In *Advances in ICT for Emerging Regions (ICTer), 2014 International Conference on*, pages 260–260. IEEE, 2014.

[12] Y. Sagi, I. Tavor, S. Hofstetter, S. Tzur-Moryosef, T. Blumenfeld-Katzir, and Y. Assaf. Learning in the fast lane: new insights into neuroplasticity. *Neuron*, 73(6):1195–1203, 2012.

[13] K. Sekihara, S. S. Nagarajan, D. Poeppel, A. Marantz, and Y. Miyashita. Reconstructing spatio-temporal activities of neural sources using an meg vector beamformer technique. *Biomedical Engineering, IEEE Transactions on*, 48(7):760–771, 2001.

[14] S. Stober, D. J. Cameron, and J. A. Grahn. Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings. In *Advances in Neural Information Processing Systems*, pages 1449–1457, 2014.

[15] L. Team. A lightweight library to build and train neural networks in theano. `https://github.com/Lasagne/Lasagne`, 2015.

[16] T. Team. A scientific computing framework for luajit. `https://github.com/torch/torch7`, 2015.