# EEG THINKER INVARIANCE IN FREQUENCY DOMAIN

*Shayan Kousha*[1,2,*], *Demetres Kostas*[1,2], *Frank Rudzicz*[1,2,3]

[1]University of Toronto; Toronto, Canada; [2]Vector Institute; Toronto, Canada ;
[3]St. Michael's Hospital; Toronto, Canada ; [*]shayan.kousha@mail.utoronto.ca

## ABSTRACT

Deep convolutional neural networks (CNN) have revolutionized machine learning techniques in computer vision and natural language processing. There is also an increasing interest among brain-computer interface (BCI) researchers to use deep CNNs and adapt successful end-to-end training techniques from computer vision and NLP. However, many of these techniques are not as successful as the researchers hoped they would be because CNNs overfit the brain signals if trained in an end-to-end manner. In this paper, we explore the idea of learning features from the frequency domain either explicitly or ideally implicitly, to make more accurate predictions using the learned features. In order to test the idea, we adapt two techniques from signal processing, namely, MFCC and SincNet introduced by Ravanelli et al. [19]. To evaluate our methods, we use the ShallowConvNet classifier introduced by Schirrmeister et al. [1] and the model we developed previously and train them all using the BCI Competition IV Dataset 2a [4]. Itll be shown in this work that explicitly learned features are not necessarily more effective than the implicitly learned ones.

## 1. INTRODUCTION

Over the past few years, deep convolutional neural networks have revolutionized almost all subfields of machine learning such as computer vision and natural language processing. However, BCI researchers have not had much success with developing models that generalize well to unseen subjects and have high accuracies as the cross-subject accuracies differ vastly from within-subject accuracies [1, 3, and 5]. Since the process of recording brain signals is challenging, most of the training dataset is not large enough to prevent the developed CNN models from overfitting, hence using very deep networks does not improve the classification rate [4]. As a result, many of the trained models learn and overfit to subject-specific features which prevent the models from generalizing to unseen subjects that are not included in the training set [1, 3, and 5]. Therefore, many of the models designed for brain signal decoding are trained and tested on the data recorded from a single subject [1, 2, 3, and 5].

However, there have been some attempts to develop techniques to remove the subject-specific features completely or reduce the effects of these features in order to improve the generalizability of the models trained on multiple subjects [6, 7]. Unfortunately, not many of these techniques use neural networks. These approaches are mainly based on regularization, user-to-user transfer, which has been used to learn CSP spatial filters, and most recently generating artificial EEG trials by relevant combinations and distortions of the original trials available in order to augment the training set size [6]. User-to-user transfer is probably the closest technique to our approach among all the mentioned approaches. However, this method relies on users for which much data is available to augment training for a target user with little data. Having more data both prevents machine learning models from fitting to subject-specific information and reduces the effects of these features.

The problem of having data sampled from different domains with unique characteristics and making predictions solely based on features that cannot discriminate between the training domains is not specific to motor imagery EEG (MI-EEG). This problem is commonly referred to as domain adaptation. Frequently used methods for transferring source include adversarial domain adaptation, which uses competing optimizations and losses for a mutual optimization [8, 9] and GAN-based domain adaptation [10, 11]. GAN-based DA usually has a generator that tries to fool the model to make the source domain look like the target one as much as possible. In this paper, we apply an adversarial domain adaptation technique called gradient reversal to one of the models in order to prevent the model from learning subject-specific features.

In this work, we also consider whether removing subject-specific information explicitly in the frequency domain is more powerful than time-domain approaches. Frequency decomposition is widely used in BCI end to end machine learning models [1, 2, 3, 4, 5, and 12]. The effectiveness of decomposing time domain signals to frequency domain has not only been shown by BCI researchers [14, 15, and 16] but also has been shown by other researchers in acoustic signals [17, 18].

We present two main models, one of which uses Mel-frequency cepstral coefficients (MFCCs) to learn the subject-specific features and the other one uses a technique called SincNet introduced by Ravanelli et al. [19]. SincNet is a

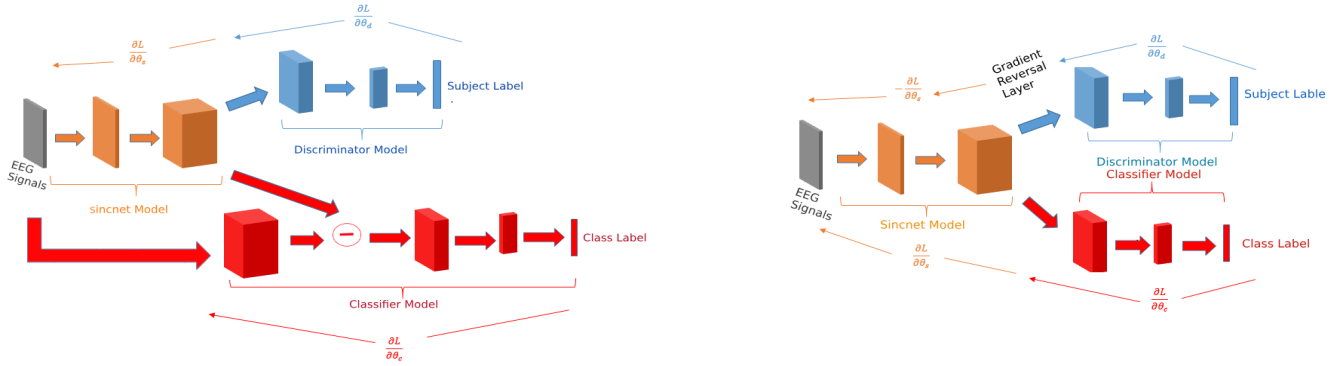


**Fig. 1**. (a) the diagram on the left displays the architecture of the Sincnet model. (b) The diagram on the right is a modified version of the Sincnet model where the gradient reversal technique is applied to the model.

smart way of doing bandpass-filtering without decomposing time domain signals. We'll discuss each approach in more details in Sec. 3.

We compare our result against a baseline model called ShallowConvNet introduced by Schirrmeister et al. [2] where they use a spatial and temporal convolutions to extract features and then perform the classification task. We also use our previously developed model as a baseline model to get a sense of the effectiveness of the techniques being discussed in this paper.

The remainder of this paper is organized as follows. After a brief review of the dataset in Sec. 2, we present in Sec. 3 the proposed methods in detail. In Sec. 4, we present the performance of our methods on the BCI IV 2A benchmark [13]. Finally, we conclude the whole paper in Sec. 5.

## 2. DATASET

We use BCI competition IV 2A dataset which is a publicly available EEG dataset [13]. It consists of 9 subjects performing four motor imagery tasks including movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Data is recorded in 6 runs separated by short breaks. One run consists of 48 trials (12 for each of the four possible classes), yielding a total of 288 trials per session. Each input data is one of these trials and includes the signals from 0.5 seconds before the start of the event to 4 seconds after the event.

We did not pre-process the signals much, except for bandpass-filtering the brain signals to 0-38 Hz and performing electrode-wise exponential moving standardization with a decay factor of 0.999 to compute exponential moving means and variances for each channel [2]. The moving means and variances were later used to standardize the signals.

In addition to bandpass-filtering, we decomposed the filtered signals to MFCCs to be used in only one of the models. The other model solely uses the filtered brain data without pre-processing the data any further. The MFCC approach computes cepstral coefficients by computing the Fourier transform of a signal and Map the powers of the spectrum obtained onto the mel scale, then transforms the output into the cepstral domain using a discrete cosine transform [15].

## 3. METHODS

The main idea that is being tested in this paper is performing the classification task in the frequency domain rather than in the time domain. The idea is to decompose the signals into their frequency components, then find a minimal set of the components that a discriminator can use to predict which subject the signals are recorded from. Later, we would remove the components of the minimal set from the pool of all components and use the remaining components to perform the motor imagery classification task. We have developed two models referred to as MFCC model and Sincnet model based on this idea.

This section first introduces the MFCC and SincNet models, then presents a short summary of the model we developed last term, and finally details training procedures and how we performed the secondary analysis.

### 3.1. MFCC Model

The input to this model is Mel-frequency cepstral coefficients (MFCCs) of EEG brain signals. MFCC is a feature extraction technique mainly applied to speech recognition task [20]. It has also been applied in EEG tasks classification in recent years and achieved high classification accuracy [21, 22]. MFCCs are a representation of the short-term power spectrum of a sound. MFCCs contain filter banks that model the ability of the human ear to resolve frequencies nonlinearly across the audio spectrum [16].

To compute MFCCs, the EEG samples were segmented into frames. The segmented frame was converted into the frequency domain using Fast Fourier Transform (FFT). MFCCs

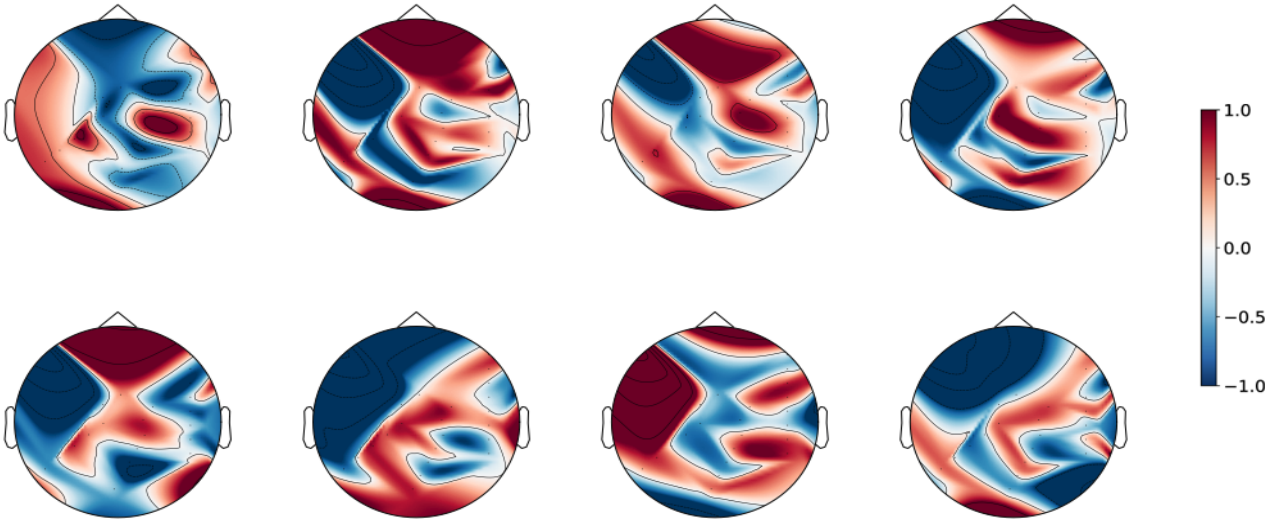


**Fig. 2**. (a) the diagram on the left displays the architecture of the Sincnet model. (b) The diagram on the right is a modified version of the Sincnet model where the gradient reversal technique is applied to the model.

are then computed by mapping the FFT spectrum onto a mel scale. The input of this model has 25 MFCC components as it results in a faster convergence when compared with a lower number of components.

The model consists of three submodels. The first submodel is learning to reweight the components in a way that helps the discriminator model to achieve higher accuracies. This submodel has a convolutional layer followed by a transposed convolutional layer. The second and third submodels are referred to as the discriminator and classifier respectfully. They both have a similar architecture with 4 convolutional layers each.

Low cross-subject classification accuracies and computationally expensive process of decomposing data to frequency domain convinced us to formulate the computations in a way that it performs the process in the time domain and implicitly extracts the same features from the frequency domain rather than explicitly performing the computation in frequency domain.

### 3.2. Sincnet Model

Similar to the MFCC model, the Sincnet model consists of three submodels (Fig. 1a). The first submodel consists of a regular convolutional layer across all the channels and a convolutional layer that performs the convolution with a predefined function g that depends on a few learnable parameters [19]. This function, inspired by standard filtering in digital signal processing, is defined such that a filter-bank composed of rectangular bandpass filters is learned. In the frequency domain, the magnitude of a generic bandpass filter can be written as the difference between two low-pass filters. After returning to the time domain using the inverse Fourier transform [23], the function g will be a function of sinc function. The learnable parameters of the function g are low and high cutoff frequencies. Therefore, this model learns different bandpass filters without decomposing the brain signals to the frequency domain.

The second and third submodels are similar to the discriminator and classifier models we developed previously with three and four convolutional layers, respectively. Exceptionally, the convolutional layers are dilated to reduce computation cost without reducing receptive field. The features learned from the first submodel are being passed to the discriminator and the classifier models. The discriminator uses these features to predict the subject and propagates the error signal back to previous layers to learn bandpass filters that help it to make better predictions. However, the classification model does not use the learned features directly. It takes in the EEG signals as well as the features, shrink the EEG signals and subtracts the features from the shrunken brain signals. This step is analogous to removing the components of the minimal set from the pool of all components in the frequency domain. Since the classifier does not use the features directly, no error signals are being propagated from this classifier to the first submodel.

We also developed another version of Sincnet model in hopes of better classification accuracies (Fig. 1 b). This version of Sincnet uses gradient reversal technique [citation] to prevent learned bandpass filters to fit unique properties of each subject. The other difference is that the classifier uses the learned features passed from the first submodel directly to perform the classification task.

| | ShallowConvNet | Last Term model | Sincnet | Sincnet + Gradient Reversal |
|---|---|---|---|---|
| Subject 1 | 53.8 | 54.86 | 49.3 | 45.13 |
| Subject 2 | 43.75 | 48.26 | 43.05 | 40.62 |
| Subject 3 | 65.62 | 64.93 | 56.94 | 54.51 |
| Subject 4 | 47.22 | 51.38 | 46.52 | 50 |
| Subject 5 | 52.43 | 56.59 | 51.73 | 49.6 |
| Subject 6 | 50 | 52.7 | 44.79 | 48.26 |
| Subject 7 | 57.63 | 60.4 | 56.94 | 47.9 |
| Subject 8 | 66.67 | 61.8 | 55.2 | 47.2 |
| Subject 9 | 65.2 | 67.36 | 63.88 | 55.9 |
| **Mean** | 55.81 | 57.58 | 52.03 | 48.79 |
| **p-value** | | 0.1548 | 0.00901 | 0.01172 |

**Table 1**. Decoding accuracy of ShallowConvNet as well as of the models we developed. The Wilcoxon signed-rank test is used to calculate the p-value

### 3.3. Compressor + Gradient Reversal

This model was developed last semester. Similar to the MFCC and Sincnet models, it consists of three parts. Part one is a compressor model that learns a smaller representation of the brain signals. A discriminator consists of a gradient reversal layer which is used to help the compressor model to remove the subject specific information from the learned smaller representation of EEG signals. The third submodel is a classifier that takes in the output of the compressor model and classifies the signals as one of the four motor imagery classed.

### 3.4. Training Procedures

Each model is tested using a held-out group of 9 subjects. The EEG signals of the remaining subjects are shuffled and split into training and validation sets with training set containing 80

### 3.5. Secondary Analysis Procedure

Explaining what a trained neural network has learned is an ongoing and challenging machine learning research area. For NNs to be commercialized and used in industry, understanding what inputs cause the trained neural network to reach a higher accuracy is critical.

To have a more meaningful comparison between the baseline model and the best hybrid model, we took an approach that generates synthetic input data. Both models are trained and tested on the same training, validation and test sets. The checkpoints of the epoch that reaches the highest validation accuracy are recorded for each model. When we are done with the training process, we set the trainability of each trainable weight and bias to false. Then a trainable input data is created and initialized randomly. Finally, the checkpoints are loaded and the models modify the trainable input data in a way that the accuracy will be close to 100% and the loss will almost be zero. Either gradient descent or the Adam optimizer can be used when performing back-propagation updates to train the inputs. Its worth noting that each model trains different input data. Results of this technique are discussed in the next section.

## 4. RESULTS

Table 1 represents the decoding accuracies of each model when trained on BCI competition IV 2A dataset. To record each of the accuracies, one of the subjects is held out, the model is trained and validated on the other subjects. Finally, the weights of the epoch with the highest validation accuracy is loaded and the model is tested on the held out subject.

The accuracy of the discriminator of MFCC model was impressive with a minimum of 85% training and validation accuracies. However, to our surprise, the classifier failed to achieve the same performance. Even though the training accuracy was high, the validation and test accuracies are in the range of 25% to 30%. Shrinking the classifier model and using regularization techniques like dropout was not effective enough to reduce the rate of overfitting significantly. As a result of surprisingly low test accuracies, the result of the MFCC model is not included in table 1.

To test significance, we choose the Wilcoxon signed-rank test and test the models against the ShallowConvNet model. This hypothesis test is used for cases where values do not follow a normal distribution. It should be noted that since multiple hypotheses are being tested, the chance of a rare event increases. Therefore, we have used the Bonferroni correction to adjust the significance level (=0.05/3=0.0167).

Even Though almost all testing accuracies of the model we developed last semester is greater than the test accuracies of ShallowConvNet, the differences between the two models are not statistically significant (p-value=0.1548). However, it should be noted that the Bonferroni correction penalizes the significance level harshly. Results are limited by the num-

ber of available subjects, and more powerful statistical tests would require a greater number of subjects.

The calculated p-values for the Sincnet model and Sincnet + Gradient Reversal are both below the adjusted significance level. The p-values let us conclude that the testing accuracies reached by these two models are statistically significantly lower than the accuracies of the baseline model.

Based on the analysis above and the results recorded in table 1, the best hybrid model is the model we developed last semester. To better compare the baseline model and this model, we used the technique explained in section 3.5. Figure 2 shows the relative intensities of the 22 EEG of synthetic data that maximized the accuracies of the models for the four motor imagery classes. There are some common patches of relative intensities between the figures demonstrating which of these channels are found to be important for the classification task by both models.

## 5. CONCOLUSION

Despite the effectiveness of frequency decomposition in BCI research [14, 15, and 16] as well as speech recognition [17, 18], our results showed that it may not aid in removing or reducing the effect of subject-specific features. Overfitting to the data is one of the reasons why this approach was not a successful one. Preventing the models from overfitting using various regularization like dropout resulted in a huge reduction of training accuracy rather than causing the validation accuracy to rise. The second reason is the slowness of the training process. The Sincnet model, which implicitly learns the features in the frequency domain, is very slow to train and computationally expensive.

It is also worth noting that having brain signals from more subjects in the dataset should improve the generalizability. Therefore, a larger dataset such as a multimodal brain-imaging dataset of simultaneous electroencephalography (EEG) and near-infrared spectroscopy (NIRS) recordings should be used to evaluate these methods even further.

## 6. REFERENCES


[1] Lawhern, Vernon J, et al. EEGNet: a Compact Convolutional Neural Network for EEG-Based BrainComputer Interfaces. Journal of Neural Engineering, vol. 15, no. 5, 2018, p. 056013.

[2] Schirrmeister, Robin Tibor, et al. Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. Human Brain Mapping, vol. 38, no. 11, 2017, pp. 53915420.

[3] Hajinoroozi, Mehdi, et al. EEG-Based Prediction of Drivers Cognitive Performance by Deep Convolutional Neural Network. Signal Processing: Image Communication, vol. 47, 2016, pp. 549555., doi:10.1016/j.image.2016.05.018.

[4] Antoniades, Andreas, et al. Deep Learning for Epileptic Intracranial EEG Data. 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), 2016, doi:10.1109/mlsp.2016.7738824.

[5] Page, Adam, et al. Wearable Seizure Detection Using Convolutional Neural Networks with Transfer Learning. 2016 IEEE International Symposium on Circuits and Systems (ISCAS), 2016, doi:10.1109/iscas.2016.7527433.

[6] Lotte, Fabien. Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based BrainComputer Interfaces. Proceedings of the IEEE, vol. 103, no. 6, 2015, pp. 871890.

[7] Waytowich, Nicholas R., et al. Spectral Transfer Learning Using Information Geometry for a User-Independent Brain-Computer Interface. Frontiers in Neuroscience, vol. 10, 2016.

[8] Ganin, Yaroslav, et al. Domain-Adversarial Training of Neural Networks. Domain Adaptation in Computer Vision Applications Advances in Computer Vision and Pattern Recognition, 2017, pp. 189209.

[9] Wulfmeier, Markus, et al. Addressing Appearance Change in Outdoor Robotics with Adversarial Domain Adaptation. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017

[10] Ghifary, Muhammad, et al. Deep Reconstruction-Classification Networks for Unsupervised Domain Adaptation. Computer Vision ECCV 2016 Lecture Notes in Computer Science, 2016, pp. 597613.

[11] Liu, et al. Coupled Generative Adversarial Networks. Curran Associates, Inc., 2016, pp. 469–477.

[12] Dose, Hauke, et al. An End-to-End Deep Learning Approach to MI-EEG Signal Classification for BCIs. Expert Systems with Applications, vol. 114, 2018, pp. 532542., doi:10.1016/j.eswa.2018.08.031.

[13] Tangermann, et al. Review of the BCI Competition IV. Frontiers, Frontiers, 30 Mar. 2012, www.frontiersin.org/articles/10.3389/fnins.2012.00055/full.

[14] https://xlescience.org/index.php/IJASIS/article/view/28/32

[15] Harati, A., et al. Improved EEG Event Classification Using Differential Energy. 2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2015, doi:10.1109/spmb.2015.7405421.

[16] Teh, Jackie, and M. P. Paulraj. Motor-Imagery Task Classification Using Mel-Cepstral and Fractal Fusion Based Features. Indian Journal of Science and Technology, vol. 8, no. 20, 2015, doi:10.17485/ijst/2015/v8i20/79066.

[17] Hertel, Lars, et al. Comparing Time and Frequency Domain for Audio Event Recognition Using Deep Learning. 2016 International Joint Conference on Neural Networks (IJCNN), 2016, doi:10.1109/ijcnn.2016.7727635.

[18] Donahue, Chris, et al. Exploring Speech Enhancement with Generative Adversarial Networks for Robust Speech Recognition. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, doi:10.1109/icassp.2018.8462581.

[19] Ravanelli, Mirco, and Yoshua Bengio. Speaker Recognition from Raw Waveform with SincNet. 2018 IEEE Spoken Language Technology Workshop (SLT), 2018, doi:10.1109/slt.2018.8639585.

[20] Muda, Lindasalwa, et al. Voice Recognition Algorithms Using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. arxiv.org/pdf/1003.4083.

[21] Yorozu Y, Hirano M, Oka K, Tagawa Y. MFCC for robust emotion detection using EEG. 2009 IEEE 9th Malaysia International Conference Communications (MICC); Kuala Lumpur. 2009 Dec 15-17. p. 98101.

[22] Abdul, W., and J.w. Wong. Cortical Activities Pattern Recognition for the Limbs Motor Action. 4th International Conference on Intelligent Environments (IE 08), 2008, doi:10.1049/cp:20081167.

[23] Rabiner, Lawrence R., and Ronald W. Schafer. Theory and Applications of Digital Speech Processing. Pearson/Prentice Hall, Theory and Applications of Digital Speech Processing, 2011