Generating EEG features from Acoustic features

Gautam Krishna
Brain Machine Interface Lab
The University of Texas at Austin
Austin, Texas

Co Tran
Brain Machine Interface Lab
The University of Texas at Austin
Austin, Texas

Mason Carnahan*
Brain Machine Interface Lab
The University of Texas at Austin
Austin, Texas

Yan Han*
Brain Machine Interface Lab
The University of Texas at Austin
Austin, Texas

Ahmed H Tewfik
Brain Machine Interface Lab
The University of Texas at Austin
Austin, Texas

arXiv:2003.00007v2 [eess.AS] 19 Mar 2020

***Abstract*—In this paper we demonstrate predicting electroencephalography (EEG) features from acoustic features using recurrent neural network (RNN) based regression model and generative adversarial network (GAN). We predict various types of EEG features from acoustic features. We compare our results with the previously studied problem on speech synthesis using EEG and our results demonstrate that EEG features can be generated from acoustic features with lower root mean square error (RMSE), normalized RMSE values compared to generating acoustic features from EEG features (ie: speech synthesis using EEG) when tested using the same data sets.**

***Index Terms*—electroencephalography (EEG), deep learning**

I. Introduction

Electroencephalography (EEG) is a non invasive way of measuring electrical activity of human brain. EEG sensors are placed on the scalp of a subject to obtain the EEG recordings. The references [1]–[3] demonstrate that EEG features can be used to perform isolated and continuous speech recognition where EEG signals recorded while subjects were speaking or listening, are translated to text using automatic speech recognition (ASR) models. In [4] authors demonstrated synthesizing speech from invasive electrocorticography (ECoG) signals using deep learning models. Similarly in [2], [5] authors demonstrated synthesizing speech from EEG signals using deep learning models. In [2], [5] authors demonstrated results using different types of EEG feature sets. Speech synthesis and speech recognition using EEG features might help people with speaking disabilities or people who are not able to speak with speech restoration.

In this paper we are interested in investigating whether it is possible to predict EEG features from acoustic features. This problem can be formulated as an inverse problem of EEG based speech synthesis. In EEG based speech synthesis, acoustic features are predicted from EEG features as demonstrated by the work explained in references [2], [5]. Predicting EEG features or signatures from unique acoustic patters might help in better understanding of how human brain process speech perception and production. Recording EEG signals in a laboratory is a time consuming and expensive process which requires the use of specialized EEG sensors and amplifiers, thus having a computer model which can generate EEG features from acoustic features might also help with speeding up the EEG data collection process as it is much easier to record speech or audio signal, especially for the task of collecting EEG data for performing speech recognition experiments.

In [6] authors demonstrated medical time series generation using conditional generative adversarial networks [7] for toy data sets. Other related work include the reference [8] where authors demonstrated generating EEG for motor task using wasserstein generative adversarial networks [9]. Similarly in [10] authors generate synthetic EEG using various generative models for the task of steady state visual evoked potential classification. In [11] authors demonstrated EEG data augmentation for the task of emotion recognition. Our work focuses only on generating EEG features from acoustic features.

We first performed experiments using the model used by authors in [5] and then we tried performing experiments using generative adversarial networks (GAN) [12]. In this work we predict various EEG feature sets introduced by authors in [2] from acoustic features extracted from the speech of the subjects as well as from acoustic features extracted from the utterances that the subjects were listening.

Our results demonstrate that predicting EEG features from acoustic features seem to be easier compared to predicting acoustic features from EEG features as the root mean square error (RMSE) values during test time were much lower for predicting EEG features from acoustic features compared to it's inverse problem when tested using the same data sets. To the best of our knowledge this is the first time predicting EEG features from acoustic features is demonstrated using deep learning models.

II. Regression and GAN model

The regression model we used in this work was very similar to the ones used by the authors in [5]. We used the exact training parameters used by authors in [5] for setting values for batch size, number of training epochs, learning rate etc. In [5] authors used only gated recurrent unit (GRU) [13]

*Equal author contribution

layers in their model but in this work we also tried performing experiments using Bi directional GRU layers where a forward GRU and backward GRU layer outputs are concatenated to produce the output of the bi directional GRU layer. The architecture of our regression model is described in Figure 1. The model takes acoustic features or mel-frequency cepstral coefficients (MFCC) of dimension 13 as input and outputs EEG features of a specific dimension at every time step. The dimension of the EEG features outputted depends on the EEG feature set used during training, as each EEG feature set had a different dimension value. The time distributed dense layer in the model has number of hidden units equal to the dimension of the EEG feature set used. The mean squared error (MSE) function was used as the regression loss function for the model. The Figure 4 shows the training convergence for the regression model when Bi directional GRU layers were used. There were two Bi-GRU layers with 256 and 128 hidden units respectively.

Generative adversarial network (GAN) [12] consists of two networks namely the generator model and the discriminator model which are trained simultaneously. The generator model learns to generate data from a latent space and the discriminator model evaluates whether the data generated by the generator is fake or is from true data distribution. The training objective of the generator is to fool the discriminator. The main motivation behind trying to perform experiments using GAN was in the case of GAN the loss function is learned where as in regression a fixed loss function (MSE) is used. However GAN models are extremely difficult to train.

Our generator model, as shown in Figure 2, consists of two layers of Bi-GRU with 256, 128 hidden units respectively in each layer followed by a time distributed dense layer with hidden units equal to the dimension of EEG feature set. During training, real MFCC features with dimension 13 from training set are fed into the generator model and the generator outputs a vector of dimension equal to EEG feature set dimension, which can be considered as fake EEG.

The discriminator model, as described in Figure 3, consists of two single layered Bi-GRU with 256, 128 hidden units connected in parallel. At each training step a pair of inputs are fed into the discriminator. The discriminator takes (real MFCC features, fake EEG) and (real MFCC features, real EEG) pairs. The outputs of the two parallel Bi-GRU's are concatenated and then fed to a GRU layer with 128 hidden units. The last time step output of the GRU layer is fed into the dense layer with sigmoid activation function.

In order to define the loss functions for both our generator and discriminator model let us first define few terms. Let P_{s_f} be the sigmoid output of the discriminator for (real MFCC features, fake EEG) input pair and let P_{s_e} be the sigmoid output of the discriminator for (real MFCC features, real EEG) input pair during training time. Then we can define the loss function of generator as $-\log(P_{s_f}) + (realEEG - fakeEEG)^2 * 0.5$ and loss function of discriminator as $-\log(P_{s_e}) - \log(1 - P_{s_f})$. The weights of Bi-GRU layers in the generator model were initialized with weights of the regression model for easier training. During test time, the trained generator model takes acoustic features or MFCC from test set as input and produces EEG features as output.

The Figure 6 shows the generator model training loss and Figure 7 shows the discriminator model training loss. The GAN model was trained for 200 epochs using adam optimizer with a batch size of 32.

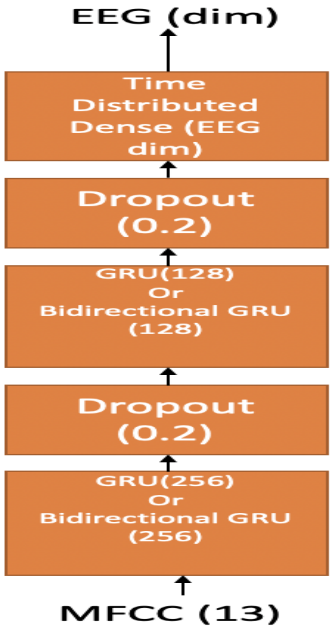

Fig. 1. Regression Model

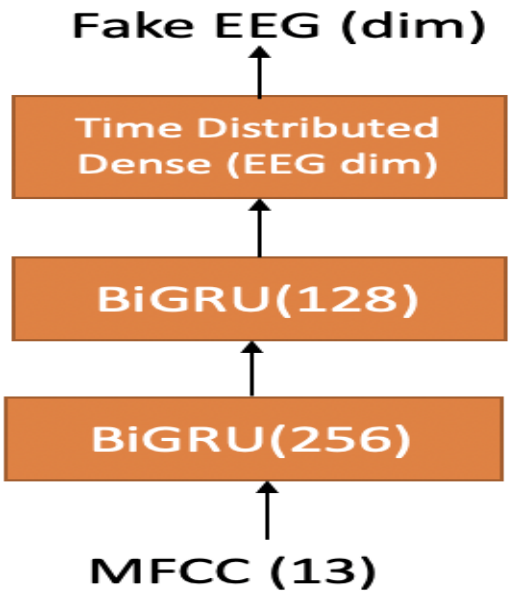

Fig. 2. Generator in GAN Model

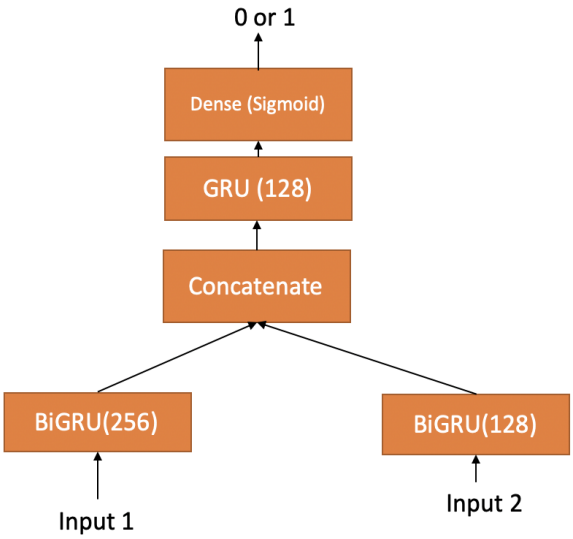

Fig. 3. Discriminator in GAN Model

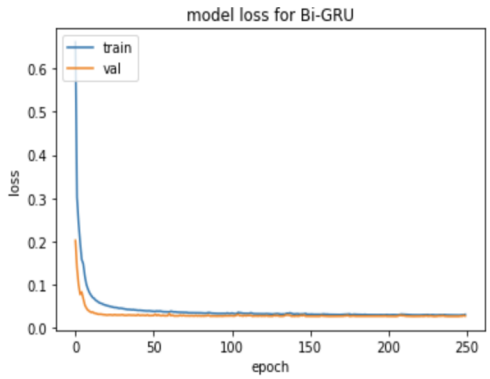

Fig. 4. Bi-GRU training loss convergence

III. Data Sets used for performing experiments

We used the data set used by authors in [5] for performing experiments. The data set contains the simultaneous speech and EEG recording for four subjects. For each subject we used 80% of the data as the training set, 10% as validation set and remaining 10% as test set. This was the main data set used in this work for comparisons. More details of the data set is covered in [5]. We will refer this data set as data set A in this paper.

We also performed some experiments using data set B used by authors in [2]. For this data set we didn't perform experiments for each subject instead we used 80% of the total data as training set, 10% as validation set and remaining 10% as test set. More details of the data set is covered in [2]. We will refer this data set as data set B in this paper. The train-test split was done randomly.

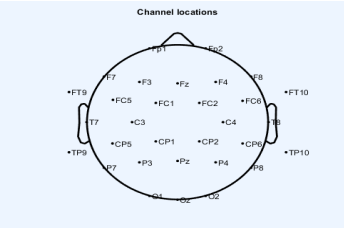

Fig. 5. EEG channel locations for the cap used in our experiments

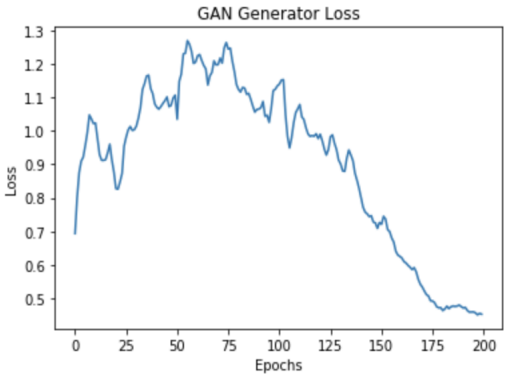

Fig. 6. Generator training loss

The EEG data used in these data sets were recorded using wet EEG electrodes. In total 32 EEG sensors were used including one electrode as ground as shown in Figure 5. The Brain Product's ActiChamp EEG amplifier was used in the experiments to collect data.

IV. EEG feature extraction details

We followed the same preprocessing methods used by authors in [1]–[3], [5] for preprocessing EEG and speech data.

EEG signals were sampled at 1000Hz and a fourth order IIR band pass filter with cut off frequencies 0.1Hz and 70Hz was applied. A notch filter with cut off frequency 60 Hz was used to remove the power line noise. The EEGlab's [14] Independent component analysis (ICA) toolbox was used to remove biological signal artifacts like electrocardiography (ECG), electromyography (EMG), electrooculography (EOG) etc from the EEG signals. We then extracted the three EEG feature sets explained by authors in [2]. The details of each EEG feature set are covered in [2]. Each EEG feature set was extracted at a sampling frequency of 100 Hz for each EEG channel [3].

The recorded speech signal was sampled at 16KHz frequency. We extracted mel-frequency cepstral coefficients (MFCC) of dimension 13 as features for speech signal. The MFCC features were also sampled at 100Hz same as the sampling frequency of EEG features.

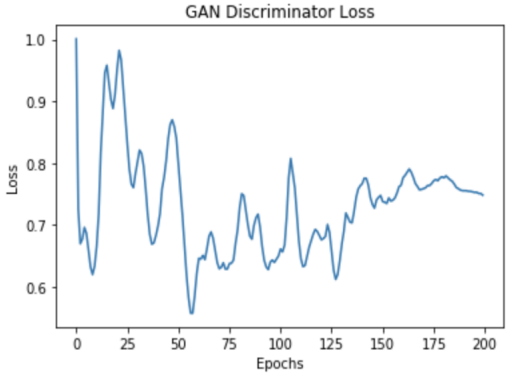

Fig. 7. Discriminator training loss

V. EEG Feature Dimension Reduction Algorithm Details

By following the dimension reduction methods used by authors in [2] we reduced EEG feature set 1 to a dimension of 30, EEG feature set 2 was reduced to a dimension of 50 using kernel principal component analysis (KPCA) [15] and EEG feature set 3 was kept at original dimension of 93. More details of explained variance plots used to identify the right feature dimensions are covered in [2].

VI. Results

We computed root mean squared error (RMSE) between the predicted EEG during test time and ground truth EEG from test set as the major performance metric to evaluate the performance of the models during test time for Data set A per subject and for Data set B.

Tables I,II,III and IV shows the results obtained for predicting various listen EEG feature sets from acoustic features for the four subjects belonging to Data set A using GRU and Bi-GRU regression models during test time. Listen EEG refers to the EEG signals recorded while subjects were listening to the utterances.

Tables V,VI,VII and VIII shows the results obtained for predicting various spoken EEG feature sets from acoustic features for the four subjects belonging to Data set A using GRU and Bi-GRU regression models during test time. Spoken EEG refers to the EEG signals recorded while subjects were speaking out loud the utterances.

We observed that RMSE values were comparable for different EEG feature sets and both GRU, Bi-GRU layers demonstrated similar results. We also computed normalized RMSE as defined by authors in [5] and observed an average normalized RMSE of **0.00068** for spoken condition for each subject and an average normalized RMSE of **0.0006** for listen condition for each subject belonging to Data set A. Our results demonstrate that the test time average RMSE and normalized RMSE values were significantly lower than values obtained by authors in [5] where they were predicting acoustic features from EEG features. These results demonstrate it is easier for a deep model to learn the mapping from acoustic features to EEG features rather than trying to learn the mapping from EEG to acoustic features.

When we performed experiments using GAN model on data set A for each subject during test time we observed an average RMSE of **0.36** for spoken, listen condition for each EEG feature set. Thus the GRU and Bi-GRU regression models outperformed GAN for predicting EEG features from acoustic features. Even though we added regularization terms to the loss function of the generator in our GAN model, it still didn't help to outperform regression models. Hypothetically GAN should have demonstrated better results than regression model as GAN also learns the loss function. Our results demonstrate the extreme difficulty of training GAN for sequence generation task. The results presented by authors in [2] also demonstrate that RNN models outperformed GAN for the task of predicting acoustic features from EEG features.

We performed experiments using GRU regression model for Data set B and observed an average RMSE of **0.23** for spoken, listen condition for each EEG feature set during test time. The observed average RMSE was again much lower compared to the RMSE values obtained by authors in [2] where they tried predicting acoustic features from EEG features using the same Data set B.

Another interesting observation we noted was that in case of the test time results demonstrated by authors in [5], the RMSE values for predicting acoustic features from EEG varied among subjects whereas in our results we observed that RMSE values during test time remained almost constant among the four subjects belonging to Data set A indicating our model was able to generalize better for all the four subjects and it also indicates the deep learning model can learn acoustic to EEG mapping easily compared to learning the mapping from EEG to acoustic features.

EEG Feature Set	**Average RMSE GRU Model**	**Average RMSE Bi-GRU Model**
Set 1	0.23	0.23
Set 2	**0.20**	0.206
Set 3	**0.19**	0.193

TABLE I
RESULTS FOR PREDICTING **LISTEN EEG FROM LISTEN MFCC FOR SUBJECT 1 DATA SET A**

EEG Feature Set	**Average RMSE GRU Model**	**Average RMSE Bi-GRU Model**
Set 1	0.22	0.22
Set 2	**0.20**	0.21
Set 3	0.19	0.19

TABLE II
RESULTS FOR PREDICTING **LISTEN EEG FROM LISTEN MFCC FOR SUBJECT 2 DATA SET A**

VII. Conclusion and Future work

In this paper we demonstrated predicting various EEG feature sets from acoustic features with very **low RMSE** and

EEG Feature Set	Average RMSE GRU Model	Average RMSE Bi-GRU Model
Set 1	0.23	0.23
Set 2	0.21	0.21
Set 3	0.19	0.19

TABLE III
RESULTS FOR PREDICTING **LISTEN EEG FROM LISTEN MFCC FOR SUBJECT 3 DATA SET A**

EEG Feature Set	Average RMSE GRU Model	Average RMSE Bi-GRU Model
Set 1	0.25	0.25
Set 2	0.23	0.23
Set 3	0.21	0.21

TABLE IV
RESULTS FOR PREDICTING **LISTEN EEG FROM LISTEN MFCC FOR SUBJECT 4 DATA SET A**

EEG Feature Set	Average RMSE GRU Model	Average RMSE Bi-GRU Model
Set 1	0.23	0.23
Set 2	0.21	0.21
Set 3	0.20	0.20

TABLE V
RESULTS FOR PREDICTING **SPOKEN EEG FROM SPOKEN MFCC FOR SUBJECT 1 DATA SET A**

EEG Feature Set	Average RMSE GRU Model	Average RMSE Bi-GRU Model
Set 1	0.23	0.23
Set 2	0.22	0.22
Set 3	0.21	0.21

TABLE VI
RESULTS FOR PREDICTING **SPOKEN EEG FROM SPOKEN MFCC FOR SUBJECT 2 DATA SET A**

EEG Feature Set	Average RMSE GRU Model	Average RMSE Bi-GRU Model
Set 1	0.24	0.24
Set 2	0.23	0.23
Set 3	0.21	0.21

TABLE VII
RESULTS FOR PREDICTING **SPOKEN EEG FROM SPOKEN MFCC FOR SUBJECT 3 DATA SET A**

EEG Feature Set	Average RMSE GRU Model	Average RMSE Bi-GRU Model
Set 1	0.24	0.24
Set 2	0.22	0.22
Set 3	0.21	0.21

TABLE VIII
RESULTS FOR PREDICTING **SPOKEN EEG FROM SPOKEN MFCC FOR SUBJECT 4 DATA SET A**

normalized RMSE values during test time. To the best of our knowledge this is the first time predicting EEG features from acoustic features is demonstrated using deep models. Our results demonstrate it is easier for a deep model to learn the mapping from acoustic to EEG features rather than trying to map the inverse.

The future work will focus on validating the results on a larger data set with more number of subjects and developing strategies to improve the training of GAN for the task of generating EEG features from acoustic features.

Our future work will also focus on using these results to better understand the underlying science behind human brain's ability to perform speech perception and production.

VIII. ACKNOWLEDGEMENT

We would like to thank Kerry Loader and Rezwanul Kabir from Dell, Austin, TX for donating us the GPU to train the models used in this work.

REFERENCES

[1] G. Krishna, C. Tran, M. Carnahan, and A. Tewfik, "Advancing speech recognition with no speech or with noisy speech," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019.

[2] G. Krishna, Y. Han, C. Tran, M. Carnahan, and A. H. Tewfik, "State-of-the-art speech recognition using eeg and towards decoding of speech spectrum from eeg," *arXiv preprint arXiv:1908.05743*, 2019.

[3] G. Krishna, C. Tran, J. Yu, and A. Tewfik, "Speech recognition with no speech or with noisy speech," in *Acoustics, Speech and Signal Processing (ICASSP), 2019 IEEE International Conference on*. IEEE, 2019.

[4] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, p. 493, 2019.

[5] G. Krishna, C. Tran, Y. Han, M. Carnahan, and A. Tewfik, "Speech synthesis using eeg," in *Acoustics, Speech and Signal Processing (ICASSP), 2020 IEEE International Conference on*. IEEE, 2020.

[6] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional gans," *arXiv preprint arXiv:1706.02633*, 2017.

[7] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[8] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, "Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals," *arXiv preprint arXiv:1806.01875*, 2018.

[9] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[10] N. K. N. Aznan, A. Atapour-Abarghouei, S. Bonner, J. D. Connolly, N. Al Moubayed, and T. P. Breckon, "Simulating brain signals: Creating synthetic eeg data via neural-based generative models for improved ssvep classification," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[11] Y. Luo and B.-L. Lu, "Eeg data augmentation for emotion recognition using a conditional wasserstein gan," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 2535–2538.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[13] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[14] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[15] S. Mika, B. Schölkopf, A. J. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel pca and de-noising in feature spaces," in *Advances in neural information processing systems*, 1999, pp. 536–542.