**Adapting an Automatic Speech Recognition System to  
Event Classification of Electroencephalograms1**

*V. Shah, R. Anstotz, I. Obeid and J. Picone*

The Neural Engineering Data Consortium, Temple University

{vinitshah, ryan.anstotz, iobeid, picone}@temple.edu

1. Research reported in this publication was most recently supported by the National Human Genome Research Institute of the National Institutes of Health under award number U01HG008468. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Identification of clinically significant events in electroencephalograms (EEGs) is a time-consuming task for neurologists [1]. EEG signals contain a variety of morphologies which relate to a combination of brain signals and noise/artifacts. Automated classification of such events has the potential to speed up the interpretation process and provide valuable input to other types of EEG decision-making software. Because of the similarities between EEGs and speech signals, both of which contain temporal/sequential information, one of our long-term goals has been to apply well-developed concepts from speech recognition to EEG processing. We have previously approached this by applying hidden Markov Models (HMMs) [2][3] using a toolkit known as HTK [4]. In this poster, we discuss the application of a new high-performance speech recognition system known as Kaldi [5] to this task. Adaptation of this technology to the EEG problem has not been as straightforward as previously thought.

Kaldi is an extremely popular open source toolkit that integrates many types of relatively new deep learning algorithms with more traditional HMM approaches. Though it is designed to be flexible, configuring it to complete non-speech recognition related tasks requires substantial modifications to the way the software handles sequential data. In this study, we adapt Kaldi to do EEG event classification on six types of EEG events: periodic lateralized epileptiform discharges (PLED), generalized periodic epileptiform discharges (GPED), spike/sharp and wave discharges (SPSW), eye movements (EYEM), artifacts (ARTF), and background (BCKG). The first three events are of clinical interest [6]. The last three events are used to model various types of background noise. We have developed a database, known as the TUH EEG Events Corpus (TUEC), that can be used to model these events [7] and have reported classification results for a number of algorithms [3]. In this study, we have developed systems based on Kaldi and compared performance to our previous approaches.

Classification is performed using a 26-dimensional feature vector consisting of Linear Frequency Cepstral Coefficient (LFCC) features which were captured from the EEG signals. The feature vector contains energy, the first seven cepstral coefficients, and the first and second derivatives of the cepstral coefficients [8]. The HMM topology for each event is the same – a 3-state Bakis model [2]. We use Gaussian Mixture Models (GMMs) for output distributions at each state in each HMM. During acoustic modeling, Kaldi HMMs are modeled based on pdf-ids [5]. Pdf-ids are GMM indices associated with individual probability density functions (PDFs). They are extracted from the context dependent decision trees where leaves of the tree represent the pdf-ids. We use 40 iterations of Viterbi training to estimate the parameters of the HMMs. A diagonal covariance matrix assumption is used at each state. Tuning experiments determined that a total 50 Gaussian mixture components were optimal.

We also evaluated the application of an adaptation technique known as Maximum Likelihood Linear Transforms (MLLTs) which performs adaptation on top of the transformed features (on pdf-ids) via Linear Discriminant Analysis (LDA). MLLT, also known as Semi-Tied Covariance (STC) [9][10], is a model-state transformation technique to estimate a global covariance matrix which allows a limited number of full covariance matrices to be shared over GMM distributions. This helps the system model correlations among features at a very low computational cost. We use 35 iterations of Viterbi training while intermittently updating the MLLT transformation matrix four times.

Finally, we developed a system based on deep neural networks [11] and HMMs, referred to as DNN-HMM, by replacing GMMs with Multi-Layer Perceptrons (MLPs) to model the observation distribution. The deep network consists of three hidden layers with 256 neurons per layer with rectified linear units (ReLU) as activation functions. The output layer contains six neurons (for each class) with Softmax activation function. The system is trained using a Stochastic Gradient Descent (SGD) optimizer and an annealing learning rate after each epoch.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Ref/Hyp** | **BCKG** | **EYEM** | **ARTF** | **PLED** | **GPED** | **SPSW** |
| **BCKG** | 71.93 | 2.59 | 7.02 | 2.28 | 7.37 | 8.81 |
| **EYEM** | 0.61 | 82.37 | 2.13 | 8.51 | 2.13 | 4.26 |
| **ARTF** | 45.19 | 2.18 | 41.24 | 2.77 | 3.81 | 4.81 |
| **PLED** | 1.85 | 4.70 | 0.70 | 54.80 | 17.62 | 20.32 |
| **GPED** | 4.85 | 2.39 | 7.46 | 20.42 | 53.32 | 11.55 |
| **SPSW** | 8.29 | 9.17 | 4.41 | 4.59 | 33.33 | 40.21 |

Table 1. Performance of the baseline GMM-HMM (HTK) system

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Ref/Hyp** | **BCKG** | **EYEM** | **ARTF** | **PLED** | **GPED** | **SPSW** |
| **BCKG** | 67.68 | 10.78 | 1.93 | 5.38 | 5.26 | 8.97 |
| **EYEM** | 42.09 | 48.90 | 5.62 | 2.80 | 0.26 | 0.33 |
| **ARTF** | 41.30 | 40.19 | 13.76 | 2.59 | 1.41 | 0.76 |
| **PLED** | 3.94 | 7.39 | 1.58 | 46.68 | 18.19 | 22.22 |
| **GPED** | 11.14 | 12.35 | 6.21 | 41.00 | 20.36 | 8.94 |
| **SPSW** | 28.62 | 16.29 | 2.95 | 6.46 | 34.26 | 11.41 |

Table 2. Performance of the baseline GMM-HMM (Kaldi) system

The Viterbi decoding algorithm [12] is used to calculate the probability of observing the sequences and output of each utterance stored in the form of a lattice [13]. Lattices contain outputs of N-best set of hypotheses of phone/word sequences. Each node in the Kaldi-lattice includes acoustic and language model scores along with time information. Since the number of events to be evaluated is only six, we kept the lattice-beam value low (0.7 – 1.0) during decoding.

Table 1 shows the performance of our baseline GMM-HMM system implemented using HTK with a total of 12,494 HMM parameters. Similarly, Table 2 provides the performance of a comparable GMM-HMM baseline implemented using Kaldi that uses 8,438 parameters. The Kaldi HMM’s Viterbi training requires ~40 minutes to train the models on 8 CPU cores whereas HTK HMMs use the Baum-Welch reestimation algorithm and require same amount of time using only 1 CPU core for training. Both these systems were scored and compared using the Epoch scoring metric [14]. Similarly, Table 3 and Table 4 compare performance of the LDA-MLLT and MLP-HMM systems, respectively. Kaldi’s LDA-MLLT system performs better than its other variants with an average detection rate of 37.42%, but still underperforms compared to HTK baseline system (57.31%). All of the Kaldi HMM variants perform very poorly on SPSW detection, since they are mainly misclassified with the GPED or BCKG events.

This study suggests that EEGs possess similar behavior to that of speech waveforms. So, speech recognition tools such as Kaldi ASR and HTK, which perform temporal/sequential classification, can be directly adapted for EEG event classification. The Kaldi HMMs developed for the six-event classification does not show any improvement in performance compared to HTK baseline system. LDA-MLLT system’s overall performance is better than its other variants but the systems, which use LDA features, perform extremely poorly on SPSW events.

References

1. K. A. Jellinger, “Niedermeyer’s Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 6th edn,” *Eur. J. Neurol.*, 2011.
2. J. Picone, “Continuous Speech Recognition Using Hidden Markov Models,” *IEEE ASSP Mag.*, vol. 7, no. 3, pp. 26–41, Jul. 1990.
3. M. Golmohammadi, A. H. H. N. Torbati, S. Lopez, I. Obeid, and J. Picone, “Automatic Analysis of EEGs Using Big Data and Hybrid Deep Learning Architectures,” *J. Clin. Neurophysiol.*, pp. 1–30, 2018.
4. “HTK,” *Machine Intelligence Laboratory, Department of Engineering, Cambridge University*, 2009. [Online]. Available: http://htk.eng.cam.ac.uk/.
5. D. Povey *et al.*, “The Kaldi speech recognition toolkit,” in *IEEE Workshop on Automatic Speech Recognition and Understanding*, 2011, pp. 1–4.
6. G. L. Krauss and R. S. Fisher, *The Johns Hopkins Atlas of Digital EEG: An Interactive Training Guide*. Johns Hopkins University Press, 2006.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Ref/Hyp** | **BCKG** | **EYEM** | **ARTF** | **PLED** | **GPED** | **SPSW** |
| **BCKG** | 62.08 | 8.69 | 15.21 | 2.29 | 1.52 | 10.21 |
| **EYEM** | 34.89 | 54.36 | 8.60 | 0.22 | 1.92 | 0.00 |
| **ARTF** | 27.03 | 38.43 | 30.67 | 0.63 | 2.52 | 0.72 |
| **PLED** | 2.26 | 11.38 | 5.31 | 43.08 | 26.87 | 11.11 |
| **GPED** | 18.71 | 3.98 | 10.58 | 29.80 | 31.01 | 5.92 |
| **SPSW** | 17.88 | 18.64 | 21.69 | 1.34 | 37.12 | 3.33 |

Table 3. Performance of an LDA-MLLT system

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Ref/Hyp** | **BCKG** | **EYEM** | **ARTF** | **PLED** | **GPED** | **SPSW** |
| **BCKG** | 73.79 | 2.71 | 13.31 | 0.35 | 2.18 | 7.63 |
| **EYEM** | 68.11 | 25.75 | 1.70 | 0.42 | 0.35 | 3.63 |
| **ARTF** | 38.65 | 27.87 | 24.35 | 0.39 | 2.58 | 6.15 |
| **PLED** | 7.70 | 3.52 | 3.96 | 46.83 | 28.38 | 9.58 |
| **GPED** | 10.80 | 0.39 | 43.78 | 21.20 | 20.87 | 2.93 |
| **SPSW** | 42.04 | 9.7 | 9.81 | 0.04 | 36.55 | 1.80 |

Table 4. Performance of a Kaldi-based DNN-HMM system

1. I. Obeid and J. Picone, “The Temple University Hospital EEG Data Corpus,” *Front. Neurosci. Sect. Neural Technol.*, vol. 10, p. 196, 2016.
2. A. Harati, M. Golmohammadi, S. Lopez, I. Obeid, and J. Picone, “Improved EEG Event Classification Using Differential Energy,” in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, 2015, pp. 1–4.
3. M. J. F. Gales, “Maximum likelihood linear transformations for HMM-based speech recognition,” *Comput. Speech Lang.*, vol. 12, no. 2, pp. 75–98, 1998.
4. M. J. F. Gales, “Semi-tied covariance matrices for Hidden Markov Models,” *Speech Audio Process. IEEE Trans.*, vol. 7, no. 3, pp. 272–281, 1999.
5. D. Povey, X. Zhang, and S. Khudanpur, “Parallel training of DNNs with Natural Gradient and Parameter Averaging,” in *International Conference on Learning Representations (ICLR)*, 2015, p. 16.
6. A. Viterbi, “Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm,” *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.
7. F. Richardson, M. Ostendorf, and J. R. Rohlicek, “Lattice-based search strategies for large vocabulary speech recognition,” in *1995 International Conference on Acoustics, Speech, and Signal Processing*, 1995, pp. 576–579.
8. V. Shah, S. Ziyabari, M. Golmohammadi, I. Obeid, and J. Picone, “Objective Evaluation Metrics for Automatic Classification of EEG Events,” *J. Neural Eng.*, pp. 1–19, 2018 (in review). *https://www.isip.piconepress.com/publications/unpublished/journals/2018/iop\_jne/metrics/*.