

Open Source Resources to Advance EEG Research

J. Smith¹, S. Jones², M. Smith¹, J. Doe³, B. Smith¹ and J. Jones¹

1. Neural Engineering Data Consortium, Temple University, Philadelphia, Pennsylvania, USA
2. Center for Corporate Research, Acme Corporation, Los Angeles, California, USA
3. Advanced Technology Laboratory, XYZ Research, New York, New York, USA
{jsmith, bsmithjjones}@temple.edu, sjones@acme.com, jdoe@xyzr.com

Since 2012, the Neural Engineering Data Consortium (NEDC) at Temple University has been providing many key data resources to support machine learning research in bioengineering [1]. In this poster, we present an update on several significant resources available from NEDC that were established to support a new generation of electroencephalogram (EEG) technology development.

Our flagship product, the Temple University Hospital Electroencephalography Corpus (TUH EEG), is the world's largest open source EEG corpus [2]. Over 1,200 sites worldwide are using this corpus to support their machine learning research. For the past 6 years we have collected every EEG session conducted at Temple University Hospital. Our historical archive goes back to 2002 and has resulted in a database of over 30,000 EEG studies from more than 16,000 patients. Each study includes a report that consists of detailed information about the patient, the patient's medical history, and a neurologist's review of the study.

Our most current release is v1.1.0 and includes 13,539 patients, 23,002 EEG sessions and reports, and 53,506 EEG files. A new feature of this release is the inclusion of three types of automatically generated annotations: (1) seizure events that include the start time, stop time, channel label and type of seizure, (2) normal/abnormal classification of a session, and (3) a six-way classification of each one-second epoch (spike/sharp waves, periodic lateralized epileptiform discharges, generalized periodic epileptiform discharges, artifacts, eye movement, and background). These annotations were generated automatically using a state-of-the-art deep learning system [3].

There are a number of subsets of this corpus that were created to support specific types of EEG analysis:

- *The TUH EEG Seizure Corpus (v1.3.0)*: supports the development of automatic seizure detection technology. The data has been manually annotated for seizure events [4]. The training dataset has been extended and contains 264 patients, 580 sessions, and 1,987 files. The development test dataset consists of 50 patients, 238 sessions, and 1,013 EEG files. Also included is a held-out blind evaluation dataset which will be used in an upcoming Kaggle-style challenge hosted by IBM. This blind evaluation set consists of 50 patients, 152 sessions, and 1,023 EEG files. More details about this challenge will follow within the next few months.
- *The TUH Abnormal EEG Corpus (v2.0.0)*: supports the development of automatic detection of abnormal EEGs. The training dataset consists of 2,310 patients, 2,717 sessions, and 2,717 EEG files. The evaluation dataset includes 253 patients, 276 sessions, and 276 EEG files. Each session is labeled as either normal or abnormal using a decision-making process described by Lopez et al. [5]. Approximately 50% of the data constitute abnormal EEG sessions.
- *The TUH EEG Slowing Corpus (v1.0.1)*: aids in the development of technology that can differentiate between post-ictal and transient slowing. Slowing can be a focal or generalized decrease in frequency and is either a part of a seizure or an independent event. This corpus consists of 38 patients, 75 sessions, and 300 annotations in 112 aggregated files. It has been used to study common error modalities in automated seizure detection [6][7].
- *The TUH EEG Epilepsy Corpus (v1.0.0)*: supports the study of the characteristics of patients diagnosed as epileptic. These patients were sorted into two classes (epileptic/non-epileptic) based on

medications listed during the recording, the clinical history of the patient, and EEG signal features associated with epilepsy. There are 237 patients, 561 sessions, and 1,648 files in this dataset.

- *The TUH EEG Events Corpus (v1.0.1)*: used to develop the six-way classification system previously described. This corpus consists of two datasets, training (359 files) and evaluation (159 files).

An important feature of all these corpora is that patient numbering is now consistent. Therefore, users can study specific patients or diseases across a broad range of conditions. Previous releases of these corpora had not been reconciled because they were preliminary releases made at different points in time and were anonymized individually.

We also have expanded the documentation about the corpus to explain electrode locations, channel labels, sample frequencies and annotation formats. There is a wealth of metadata now available for each patient and session.

In addition to data resources, NEDC also provides some supporting tools to facilitate research. These include:

- *The NEDC Demo System (v0.4.1)*: a visualization tool developed to rapidly annotate EEG signals [8]. Our demo system allows users to annotate EEG signals on a per-channel basis. This tool is written in Python, using the PyQt toolkit, and is easily customized to support specific annotation tasks. We have used it on most popular operating systems including Windows, Linux, and Mac systems. A cohort retrieval system has been integrated into this viewer that allows users to query the TUH EEG Corpus for sessions that match particular search criteria. Both signal and report events can be searched.
- *NEDC Eval EEG (v1.2.0)*: a standardized scoring package that is an important piece of any common evaluation framework [9]. This software implements a variety of popular scoring metrics based on common measures such as sensitivity, specificity, and Cohen's kappa statistic. It also includes new metrics based on time-aligned and epoch scoring, providing a more balanced view of performance.

In this poster, we will discuss these resources and describe how they fit together to promote the development of deep learning technology for automatic interpretation of EEGs. All of the data and resources presented here are freely available at https://www.isip.piconepress.com/projects/tuh_eeg/downloads/. No licensing or data sharing agreements are needed. An automated registration process provides users with a username and password to access the data. The unencumbered nature of these resources is a very important differentiating feature of these NEDC resources.

ACKNOWLEDGMENTS

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. This material is also based in part upon work supported by the National Science Foundation under Grant No. IIP-1622765. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] L. Veloso, J. R. McHugh, E. von Weltin, I. Obeid, and J. Picone, "Big Data Resources for EEGs: Enabling Deep Learning Research," presented at the *IEEE Signal Processing in Medicine and Biology Symposium*, 2017, p. 1.

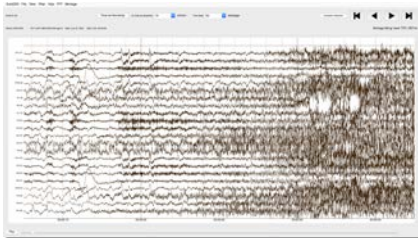
- [2] I. Obeid and J. Picone, "The Temple University Hospital EEG Data Corpus," *Front. Neurosci. Sect. Neural Technol.*, vol. 10, p. 00196, 2016.
- [3] M. Golmohammadi, A. Harati Nejad Torbati, S. Lopez, I. Obeid, and J. Picone, "Automatic Analysis of EEGs Using Big Data and Hybrid Deep Learning Architectures," *Front. Hum. Neurosci.*, pp. 1–30, 2018 (in review). https://www.isip.piconepress.com/publications/unpublished/journals/2018/frontiers_neuroscience/hybrid/.
- [4] V. Shah, E. von Weltin, S. Lopez, J. R. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, and J. Picone, "The TUH EEG Seizure Corpus," *Front. Neurosci.*, pp. 1–9, 2018 (in review). https://www.isip.piconepress.com/publications/unpublished/journals/2018/frontiers_neuroinformatics/tuh_eeg_seizure/.
- [5] S. Lopez, I. Obeid, and J. Picone, "Automated Interpretation of Abnormal Adult Electroencephalograms," presented at the 26th Conference on Intelligent Systems for Molecular Biology, 2018, p. 1.
- [6] M. Golmohammadi, I. Obeid, and J. Picone, "Deep Residual Learning for Automatic Seizure Detection," presented at the 26th Conference on Intelligent Systems for Molecular Biology, 2018, p. 1.
- [7] E. von Weltin, T. Ahsan, V. Shah, D. Jamshed, M. Golmohammadi, I. Obeid, and J. Picone, "Electroencephalographic Slowing: A Primary Source of Error in Automatic Seizure Detection," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, 2017, pp. 1–5.
- [8] N. Capp, E. Krome, I. Obeid, and J. Picone, "Rapid Annotation of Seizure Events Using an Extensible Visualization Tool," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, 2017, p. 1.
- [9] 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/.

Abstract

- An electroencephalogram (EEG) is a multi-channel signal which describes the electrical activity in the brain through electrodes placed on the scalp.
- Other EEG visualization tools do not allow users to annotate directly over their signal displays. Similarly, it is difficult to add new visualizations to these tools.
- This visualization tool allows users to create annotations through the interaction of context menus and click and drag actions.
- A search API has been integrated to this viewer. Users can query the TUH EEG Corpus for reports and EEG signals.
- The features within the viewer are customizable through a preferences window, allowing users to save settings based on their annotation and visualization needs.
- This visualization tool is written in PyQt, which allows for Graphical User Interface programming that is supported across multiple platforms.

Overview of the Visualization Software

- The viewer loads EEG signals in a European Data Format (EDF), and provides a user interface to view and manipulate the signal.



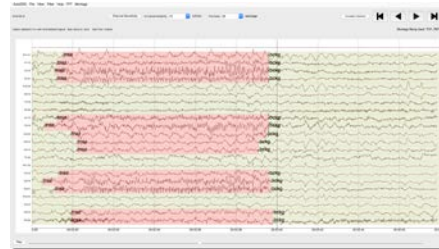
Clinical EEG Data

- EEG signals are stored as sampled data signals in an EDF format.
- To reduce the noise and enhance certain features of the signal, a montage is applied.
- A montage is used to redefine channels as the difference of channel voltages. This kind of differential analysis reduces noise.
- A TCP montage is most commonly used. However the tool offers a user interface to customize and create any montage to match the given EDF file.



Facilitating EEG Annotation

- The visualization tool provides an interface to view, create, and edit annotations directly over the EEG signal.
- Annotations are colored overlays containing a label defining the event type. These events include various seizures, artifacts, and other markings used for our research.
- An annotation is created independently of the other channels, allowing for the precise annotating of boundaries and event types.



- The tool utilizes annotation navigation buttons to easily scroll through the EEG to find regions of interest.
- Creating annotations is accomplished through selecting channels and the annotation label via a windowed interface.



Increasing Visibility

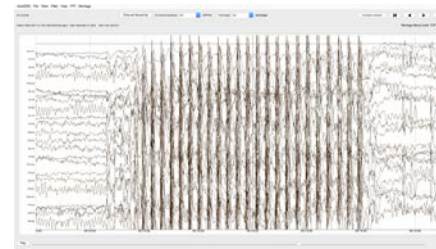
- The ability to change the timescale of the plotting window from the traditional 10 seconds allows for users to more easily visualize and annotate events on a single window.
- Channels suffering from heavy lead artifact or containing electrocardiogram (EKG) signals often clutter the plotting visualization. A per-channel gain setting is utilized to increase the visibility of events during such cases.
- A customizable filter is also integrated into signal plotting, allowing the user to remove various frequency ranges and bands. This filter includes various bandpass presets to, for example, only show frequencies in the alpha or theta bands. A common use of this filter is to apply a band-reject filter to remove the 60Hz hum from the EEG signal.

Alternate Visualizations

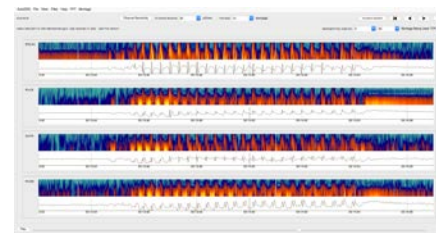
- The visualization tool currently supports three visualizations: Waveform (default), Spectrogram, and Energy.
- The Waveform view is the format through which professional neurologists traditionally prefer to view EEG data.
- The Spectrogram view is a graphical plot of the frequency content of the EEG signal as a function of time. The data is plotted with time on the horizontal axis, frequency on the vertical axis, and color used to indicate the intensity of the frequency.
- The Energy view calculates the energy of the signal over time, and plots the data on a logarithmic or linear scaled axis.

Sample Visualizations

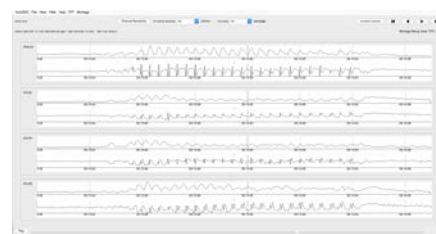
- Waveform Only:



- Waveform/Spectrogram Combined:



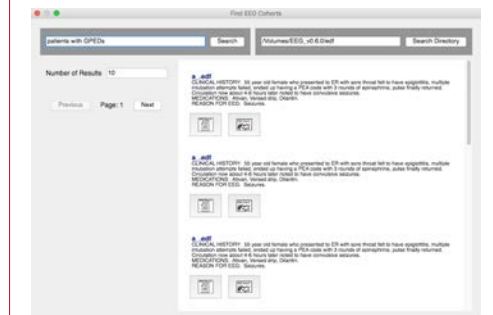
- Waveform/Energy Combined:



- Above is an absence seizure event, shown through each visualization supported by the software. Each visualization can be uniquely utilized to extract information from an EEG event.

Cohort Retrieval

- A cohort retrieval system allows users to find EEGs similar to the EEG being viewed.
- Keyword search and natural language interfaces are supported (e.g., "show me all the EEGs that have...").
- The search API queries the TUH EEG Database, and returns EEG reports matching the search query.
- Cohort retrieval integrates knowledge from the unstructured EEG reports with EEG events automatically extracted from the signal.
- The cohort retrieval system returns a list of top ranked reports and EEG signals based on the query.



- This interface establishes the ability to view multiple windows of EEG signals, along with their respective medical reports.
- Each new plotting window contains all functionality provided in the framework that the original visualization tool utilizes.

Summary

- This visualization tool facilitates EEG annotation through a simple and efficient user interface.
- The viewer has multiple visualization options, allowing users to analyze EEG events in manners outside the traditional waveform visualization.
- It has been programmed in a fashion such that new features and visualizations can easily be integrated in the existing framework.
- All features and options contained within the tool are customizable through an integrated preferences window.
- Since the tool is written in PyQt, it runs on any platform that supports Python.

Acknowledgements

- 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.