Annotation of Ambulatory EEGs

A. Melles[1], M. Paderewski[1], R. Oymann[1], V. Shah[1], J. Salazar[2], I. Obeid[1] and J. Picone[1]

1. The Neural Engineering Data Consortium, Temple University, Philadelphia, Pennsylvania, USA
2. Natus Medical Inc., Middleton, Wisconsin, USA
{anne-mai.melles, makeda.paderewski, renee.oymann, iobeid, picone}@temple.edu
{vinit.shahj@gmail.com, Jesse.Salazar@natus.com}

The Neural Engineering Data Consortium (NEDC) began annotating large amounts of EEG data in 2012 [1][2]. The most current release of the Temple University EEG Corpus, TUEG v2.0.1, consists of 26,846 sessions from 14,987 patients. TUEG has become one of most significant open-source resources available in the community. Over 10,000 researchers have subscribed to this corpus. However, the TUEG data [3] consists of pruned EEGs [4][5]. In clinical settings, technicians condense long term studies highlight any potential abnormalities, reducing the burden of reading long-term EEGs, and allowing a neurologist to focus on diagnosing the patient accurately and more efficiently. This results in data that has been split into a series of shorter files, destroying the continuous nature of the data. Gaps between files have been discarded (historically to save disk space and reduce review time), which prevents reconstruction of the original continuous recording. This makes it difficult to use this data to develop seizure prediction algorithms, accurately measure false alarm rate, or assess robustness to real-world artifacts like patient and electrode movement.

Natus Medical Inc. has collected a large corpus of ambulatory EEG data (NMAE). Ambulatory EEG data is collected from a patient by using a portable EEG device that continuously monitors and records brain activity while patients go about their daily activities. The dataset, which consists of over 1,400 studies containing 15,529 uniformly annotated seizure onset annotations marked as "sz onset," was analyzed to identify and select studies of interest for collaborative scoring with the NEDC EEG annotation team. This type of data will support a wide range of research and technological developments, including seizure prediction, long-term contextual modeling, artifact detection, and adaptation to the ambient environment and patient. In this abstract, we discuss our approach to manually annotating this data.

Ambulatory EEG data captures brain activity in real-world, everyday settings, reflecting natural variations and artifacts like movement, making it more representative of typical patient behavior. Unlike stationary EEG data, which is collected in controlled, clinical environments with minimal external influences, ambulatory EEG data is inherently noisier and lacks an EKG channel. However, the extended duration of the recordings often results in a substantial amount of sleep data, providing valuable insights into brain activity over long periods, including during sleep. This also allows researchers to capture seizures and related events, such as absence seizures and Brief Ictal Rhythmic Discharges (BIRDs) [6].

TUEG and NMAE differ in several important aspects. TUEG contains pruned recordings of standard clinical records with an average file duration of 23.3 mins. These are primarily sampled at 250 Hz, though there are also a range of sample frequencies used. Over 40 different channel configurations are included in the corpus. There are an average of 31 EEG-specific channels supplemented with additional channels for bursts, EKG, EMG, and photic stimuli [3]. In contrast, NMAE contains recordings totaling 72 hours per patient that are sampled at either 200 Hz or 256 Hz. They use a bipolar montage and do not contain EKG channels. TUEG can be clustered into four different montages: (1) the most popular bipolar montage is the Temporal Central Parasagittal (01_tcp_ar), (2) a Linked Ears Reference (02_tcp_le) montage, (3) a 20-channel Averaged Reference (03_tcp_ar_a) montage, and (4) a 20-channel Linked Ears Reference (04_tcp_le_a) montage. Details on the electrode configurations and recording conditions can be found in [7]. NMAE recordings use different channel labels but essentially follow the 01_tcp_ar montage. This makes it easy to run experiments on both corpora simultaneously.

TUEG benefited from access to detailed clinical reports, which allowed for more precise annotations. For example, generalized seizures were classified by type (e.g., myoclonic, atonic, tonic, and clonic). NMAE is being annotated in two passes. In the first pass, we are focused on the annotation of seizure events, specifically categorizing them into focal and generalized seizures. At this stage, we do not differentiate between the subtypes of focal seizures. Thus, we do not separately identify complex partial seizures (focal unaware seizures) or simple partial seizures (focal aware seizures). In the context of generalized seizures, our annotation process includes a specific designation for absence seizures, while all other types of generalized seizures are uniformly categorized under a general classification of generalized seizures. This approach is adopted because, in the absence of clinical reports, it is challenging to accurately determine the specific type of focal or generalized seizure.

In the second pass, we are annotating Rhythmic Periodic Patterns (RPPs) and Brief Ictal Rhythmic Discharges (BIRDs). Annotating RPPs presents a significant challenge due to the absence of EKG channels, complicating the differentiation between rhythmic and periodic brain-related activity and pulse artifacts. According to the new ACNS guidelines [6], RPPs, including Generalized Periodic Discharges (GPD), Lateralized Periodic Discharges (LPD), and Rhythmic Delta Activity (RDA), indicate epilepsy without ongoing seizures.

We annotate the data using our tool, nedc_eas, as described by Capp et al. [8]. This tool addresses limitations of existing EEG visualization tools by allowing us to open EDF files in the annotation tool, display annotations in a time-aligned format, directly manipulate annotations, and create a CSV file for further analysis. An annotator can typically process approximately 10 files, corresponding to 5 hours of EEG recordings, in a single hour. To our pleasant surprise, the data is surprisingly clean. Although we considered employing various noise reduction techniques, prior informal testing in our lab demonstrated minimal improvement to the annotation process using data processed through noise reduction algorithms. Further, and equally important, the performance of our baseline seizure detection system was not significantly improved.

Annotations are reviewed by a project manager, who selects data for review based on observed patterns and knowledge of each annotator's behavior. The team also meets weekly to discuss challenging cases encountered in the files and arrive at a committee consensus. We operate in a discussion-based environment, collaboratively researching and determining the best approaches to handle specific activities. To enhance efficiency, we use our real-time seizure detection system [9] to triage the data. This system identifies segments with potential seizure activity, allowing annotators to focus on these relevant sections rather than spending time on non-relevant data. Our goal is to train the model on NMAE data and compare the outcomes with TUEG data.

This new corpus will lay the foundation for a new generation of seizure prediction technology and allow exploration of how we can integrate information about BIRDs and RPPs to improve performance. It will support experiments with the robustness of systems trained under mismatched training conditions and allow accurate assessment of false alarm rates (FAs). A low false alarm rate (e.g., 1 FA per 24 hours of data) is one of the most important benchmarks for the development of clinically acceptable technology. Discussions are ongoing about the release of this corpus as open source data.

REFERENCES

[1]   C. Ward, I. Obeid, J. Picone, and M. Jacobson, "Leveraging Big Data Resources for Automatic Interpretation of EEGs," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (SPMB), New York City, New York, USA, 2013, p. 1. doi: *http://www.isip.piconepress.com/ publications/conference_presentations/2013/ieee_spmb/eeg/.*

[2]   G. Buckwalter, S. Chhin, S. Rahman, I. Obeid, and J. Picone, "Recent Advances in the TUH EEG Corpus: Improving the Interrater Agreement for Artifacts and Epileptiform Events," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (SPMB), I. Obeid, I. Selesnick, and J. Picone, Eds., Philadelphia, Pennsylvania, USA: IEEE, 2021, pp. 1–3. doi: *10.1109/SPMB52430.2021.9672302.*

[3]   I. Obeid and J. Picone, "The Temple University Hospital EEG Data Corpus," in *Augmentation of Brain Function: Facts, Fiction and Controversy. Volume I: Brain-Machine Interfaces*, 1st ed., vol. 10, M. A. Lebedev, Ed., Lausanne, Switzerland: Frontiers Media S.A., 2016, pp. 394–398. doi: *10.3389/fnins.2016.00196.*

[4]   Natus Medical, "Natus EEG/PSG Tutorial: Neuroworks/SleepWorks 9 - How to Clip & Prune." Accessed: Jul. 26, 2024. [Online]. url: *https://www.youtube.com/watch?v=fUSjWyw9BYA.*

[5]   P. Fish, "Making Sense of EEG Waves: Pruning the Data." Accessed: Jul. 26, 2024. [Online]. url: *https://stratusneuro.com/making-sense-of-eeg-waves-pruning-the-data/.*

[6]   L. J. Hirsch et al., "American Clinical Neurophysiology Society's Standardized Critical Care EEG Terminology: 2021 Version.," *J Clin Neurophysiol*, vol. 38, no. 1, pp. 1–29, Jan. 2021, doi: *10.1097/WNP.0000000000000806.*

[7]   S. Ferrell *et al.*, "The Temple University Hospital EEG Corpus: Electrode Location and Channel Labels," Temple University, Philadelphia, Pennsylvania, USA, 2020. doi: *https://www.isip. piconepress.com/publications/reports/2020/tuh_eeg/electrodes.*

[8]   N. Capp, E. Krome, I. Obeid, and J. Picone, "Facilitating the Annotation of Seizure Events Through an Extensible Visualization Tool," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (SPMB), I. Obeid, I. Selesnick, and J. Picone, Eds., Philadelphia, Pennsylvania, USA: IEEE, 2017, p. 1. doi: *10.1109/SPMB.2017.8257043.*

[9]   V. Khalkhali, N. Shawki, V. Shah, M. Golmohammadi, I. Obeid, and J. Picone, "Low Latency Real-Time Seizure Detection Using Transfer Deep Learning," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (SPMB), I. Obeid, I. Selesnick, and J. Picone, Eds., Philadelphia, Pennsylvania, USA: IEEE, 2021, pp. 1–7. doi: *10.1109/SPMB52430.2021.9672285.*