# **Annotation of Ambulatory EEGs**

A. Melles<sup>1</sup>, M. Paderewski<sup>1</sup>, R. Oymann<sup>1</sup>, V. Shah<sup>1</sup>, J. Salazar<sup>2</sup>, I. Obeid<sup>1</sup> and J. Picone<sup>1</sup>

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.

# ACKNOWLEDGEMENTS

The material presented in this abstract was supported by Natus Medical Incorporated. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of these sponsors.

# 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.
- 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.000000000000806.
- [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.



# Annotation of Ambulatory EEGs

A. Melles, M. Paderewski, R. Oymann, V. Shah, I. Obeid, and J. Picone The Neural Engineering Data Consortium, Temple University



- The Neural Engineering Data Consortium (NEDC) began annotating large amounts of EEG data in 2012.
- TUEG v2.0.1. consists of 26.846 sessions from 14.987 patients, and has over 10,000 subscribers.
- TUEG, however, are pruned recordings in which gaps have been discarded, preventing the development of technology requiring continuous recordings.
- Natus Medical Inc. has collected continuous ambulatory EEG data - over 1.400 studies consisting of 72-hour recordings. We refer to this as the Natus Medical Ambulatory EEG (NMAE) Corpus.
- We are annotating NMAE according ACNS 2021 guidelines that emphasize new events such as rhythmic and periodic patterns (RPPs) and brief potentially ictal rhythmic discharges (BIRDs).
- The Natus Medical Ambulatory EEG (NMAE) Corpus will support the development of a new generation of seizure prediction technology and accurate assessment of false alarm rates for existing systems.

#### **Continuous Ambulatory EEGs**

· Ambulatory EEG data is collected using portable equipment while patients go about their daily activities.



- Stationary EEG data is collected in a clinical setting under controlled conditions.
- The absence of an EKG channel makes it hard to distinguish between EKG artifacts (often found on A1
- and A2 channels) and brain activity. Tomal & Contraction of the Sol
- · The lack of clinical reports complicates identifying the type of seizure.
- · This also allows researchers to capture seizures and related events, such as absence seizures and BIRDs that are hard to spot in stationary settings.



#### The NMAE Corpus

- 1,400 studies containing 72 hours of data per subject.
- · 15,529 uniformly annotated seizure onset annotations marked as "sz onset."



- The data follows our standard Temporal Central Parasagittal layout in a 10-20 system (01\_tcp\_ar bipolar montage).
- There are two different sample frequencies 200 Hz (86%) and 250 Hz (14%).
- · There are 35 unique channel labels with 6 electrode configurations:
- Base labels that are consistent across all configurations: Frantal ED4 ED2 ED7 (association as Deviately D2 D4 D7

missing), F7, Temporal: T3, Central: C3, C Some unio	F8, F3, F4, FZ F8, F3, F4, FZ , T4, T5, T6 ¼, CZ que channel configur	<ul> <li>Occipital: 01, 02</li> <li>Ears: A1, A2</li> <li>Other: X1, X2, EDF ANNOTATIONS ations:</li> </ul>
Channel Configs	Additions	Channel Configurations
25	None	9000 8000 <b>891</b> 3
26	DIF1	7000 -
27	OSAT. PR	6000

studies of interest for collaborative scoring.									
These s	eizure	annotation	s were	us	sed	to	dow	n-s	elect
35	FPZ			25	26	27	28	34	35
			0 .		_	1000			

- The initial set of studies selected for annotation were selected as studies with the highest identified seizure counts - we refer to this as high-yield data.
- · We have annotated 3,503 files (30 studies):
- □ 2,828 contained only background
- □ 675 contained seizures: 52 focal, 200 generalized, 434 absence

28

DIF2

DIF3, DIF4, DC1, DC2, DC3, DC4

- TUSZ Seizure Types – TUSZ Seizure Types – NMAE spsz bckg cpsz tirda seiz tnsz frida fnsz cnsz grda gnsz tcsz lpd absz gped gpd intr pled bird • fn sz • a bsz • sp sz e g rsz ■ g nsz ■ fn sz = a bsz
- TUSZ (v2.0.3) consists of 8,239 files;:
- □ 1,391 contained seizures: 805 focal, 439 generalized, 18 absence, 5 simple partial, 168 complex partial, 11 tonic, 22 tonic-clonic
- The differences between TUSZ and NMAE underscores the importance of collecting ambulatory EEG data since it increases the likelihood of observing absence seizures.

## J. Salazar Natus Medical Incorporated

### **The Annotation Pipeline**

- · We are developing two documents that track our decisions about how to annotate the data:
- □ Reports where we document all the decisions we have made and notes about RPPs for future work.
- Interesting activities that could be confused for ictal activity.

#### Section of the reports document:

Patient number: 6705f62cc8919954be3ef9d0f93c0ecb Seizures: generalized, activities that come across as focal seizure, however, waxing and waning - not prominent enough: decided not to

RPPs: LPDs and GPDs can be confused as focal seizure but frequency is < 2.5 Hz

Patient number: 7d109c9692eca076f3b5813dc7ded0ca Seizures: generalized seizures with long post-ictal phase that includes GPDs- chose to annotate 2 seconds in the post-ictal phase because otherwise too long. Annotated more prominent seizures, onset seems to be long. Looked into the activity with unipolar montage and it cleared out that there are no focal seizures but everything is generalized. RPPs: GPDs in post-ictal phase

#### · Some examples of challenging waveforms:

□ Shivering: usually occurs with underlying ictal activity but in this case it had no spike and wave activity that could indicate a seizure.

nn nn bb	 iici ni chaola.	Minere		Mahi
13-01 192-18 18-14 14-16	 deten and office	Wang da	Ì	
		navieset Norice	Anna Alexandre	
14-AL 19-11 14-AL 14			-	
102-14 14-03 (44) 14-03		(1.). 		

#### A closer examination reveals:

□ Seizures of patients with Lennox-Gastaut syndrome (LGS) are difficult to spot since the background is slow spike and wave activity with intermittent bursts of sharp spikes.





· The snapshot above shows a generalized seizure in an LGS patient, while the image below shows the normal background EEG pattern for LGS.

#### **Progress To Date**

- · An annotator can typically process approximately 2.5 hours of data in one hour (.4xRT).
- Annotations are reviewed by a project manager, who selects data for review based on observed patterns.
- The team also meets weekly to discuss challenging cases and arrive at a committee consensus. Each patient is reviewed in detail, with 3 to 8 files typically needed to make a decision out of the 132 total files.
- This far we have annotated ~2.500 hours of data.
- · We triage the data with an EEG system based on a ResNet18 architecture:



- This system has a low raise alarm rate (0.70 PAS/24F) and a reasonable sensitivity (42.05%).
- This is a real-time system that has approximately 120 secs of latency.
- It was tuned to have a higher FA rate but a low false
- negative rate. · We are in the process of
- adapting this system to NMAE to assess the improvement in performance achieved by adaptation and to compare performance on NMAE to TUSZ.

#### Summary and Future Work

- Annotated ambulatory EEG data represents a great opportunity to improve seizure prediction and seizure detection technology, and to better explain our what our models have learned.
- Future work will involve annotating RPPs and BIRDs, which will pave the way for a new generation of software that provides neurologists with much more relevant information about a patient's health.
- The development of machine learning technology to detect RPPs and BIRDs in addition to seizures represents an exciting new research direction that will be enabled by this work.
- We will also be able to explore how we can simultaneously leverage TUEG and NMAE to develop more robust seizure detection technology.

#### Acknowledgments

The material presented in this abstract was supported by Natus Medical Incorporated. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of our sponsors.

> To learn more about this project, please use this URL: https://isip.piconepress.com/projects/natus eeg/



