Recent Advances in the TUH EEG Corpus: Improving the Interrater Agreement for Artifacts and Epileptiform Events

G. Buckwalter, S. Chhin, S. Rahman, I. Obeid and J. Picone

The Neural Engineering Data Consortium, Temple University, Philadelphia, Pennsylvania, USA
{grace.buckwalter, sidney.chhin, rahman, iobeid, picone}@temple.edu

The Temple University Hospital EEG Corpus (TUEG) [1] is the largest publicly available EEG corpus of its type and currently has over 5,000 subscribers (we currently average 35 new subscribers a week). Several valuable subsets of this corpus have been developed including the Temple University Hospital EEG Seizure Corpus (TUSZ) [2] and the Temple University Hospital EEG Artifact Corpus (TUAR) [3]. TUSZ contains manually annotated seizure events and has been widely used to develop seizure detection and prediction technology [4]. TUAR contains manually annotated artifacts and has been used to improve machine learning performance on seizure detection tasks [5]. In this poster, we will discuss recent improvements made to both corpora that are creating opportunities to improve machine learning performance.

Two major concerns that were raised when v1.5.2 of TUSZ was released for the Neureka 2020 Epilepsy Challenge were: (1) the subjects contained in the training, development (validation) and blind evaluation sets were not mutually exclusive, and (2) high frequency seizures were not accurately annotated in all files. Regarding (1), there were 50 subjects in dev, 50 subjects in eval, and 592 subjects in train. There was one subject common to dev and eval, five subjects common to dev and train, and 13 subjects common between eval and train. Though this does not substantially influence performance for the current generation of technology, it could be a problem down the line as technology improves. Therefore, we have rebuilt the partitions of the data so that this overlap was removed. This required augmenting the evaluation and development data sets with new subjects that had not been previously annotated so that the size of these subsets remained approximately the same. Since these annotations were done by a new group of annotators, special care was taken to make sure the new annotators followed the same practices as the previous generations of annotators. Part of our quality control process was to have the new annotators review all previous annotations. This rigorous training coupled with a strict quality control process where annotators review a significant amount of each other's work ensured that there is high interrater agreement between the two groups (kappa statistic greater than 0.8) [6].

In the process of reviewing this data, we also decided to split long files into a series of smaller segments to facilitate processing of the data. Some subscribers found it difficult to process long files using Python code, which tends to be very memory intensive. We also found it inefficient to manipulate these long files in our annotation tool. In this release, the maximum duration of any single file is limited to 60 mins. This increased the number of edf files in the dev set from 1012 to 1832.

Regarding (2), as part of discussions of several issues raised by a few subscribers, we discovered some files only had low frequency epileptiform events annotated (defined as events that ranged in frequency from 2.5 Hz to 3 Hz), while others had events annotated that contained significant frequency content above 3 Hz. Though there were not many files that had this type of activity, it was enough of a concern to necessitate reviewing the entire corpus. An example of an epileptiform seizure event with frequency content higher than 3 Hz is shown in Figure 1. Annotating these additional events slightly increased the number of seizure events. In v1.5.2, there were 673 seizures, while in v1.5.3 there are 1239 events.

One of the fertile areas for technology improvements is artifact reduction. Artifacts and slowing constitute the two major error modalities in seizure detection [3]. This was a major reason we developed TUAR. It can be used to evaluate artifact detection and suppression technology as well as multimodal background models that explicitly model artifacts. An issue with TUAR was the practicality of the annotation tags used

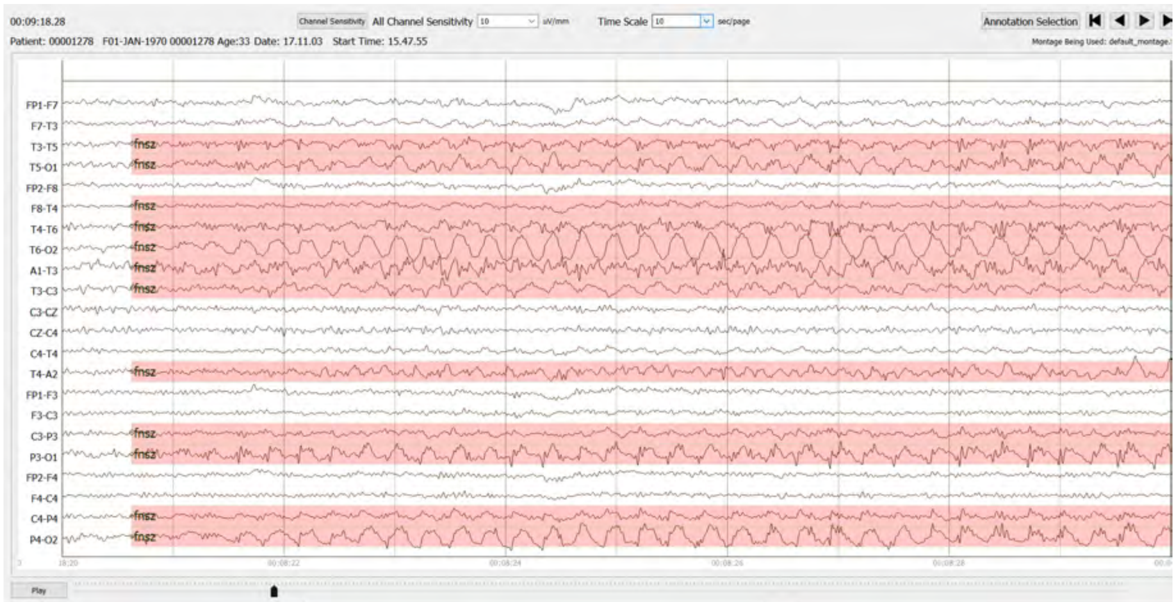


Figure 1. An example of an epileptiform seizure event with frequency content greater than 3 Hz

when there are multiple simultaneous events. An example of such an event is shown in Figure 2. In this section of the file, there is an overlap of eye movement, electrode artifact, and muscle artifact events. We previously annotated such events using a convention that included annotating background along with any artifact that is present. The artifacts present would either be annotated with a single tag (e.g., MUSC) or a coupled artifact tag (e.g., MUSC+ELEC). When multiple channels have background, the tags become crowded and difficult to identify. This is one reason we now support a hierarchical annotation format using XML – annotations can be arbitrarily complex and support overlaps in time.

Our annotators also reviewed specific eye movement artifacts (e.g., eye flutter, eyeblinks). Eye movements are often mistaken as seizures due to their similar morphology [7][8]. We have improved our understanding of ocular events and it has allowed us to annotate artifacts in the corpus more carefully.

In this poster, we will present statistics on the newest releases of these corpora and discuss the impact these improvements have had on machine learning research. We will compare TUSZ v1.5.3 and TUAR v2.0.0

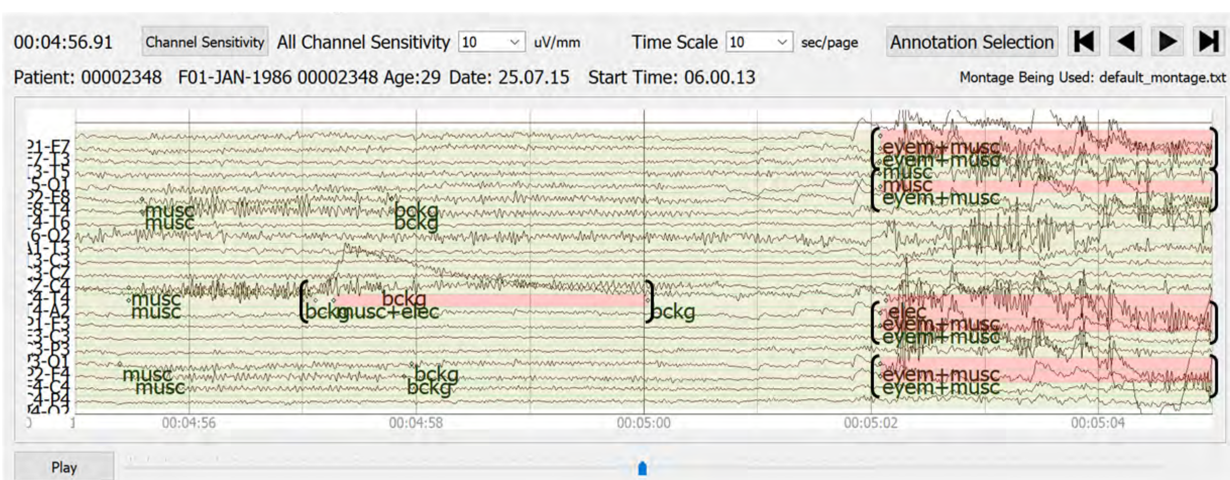


Figure 2. An example of an EEG file with multiple events occurring simultaneously

with previous versions of these corpora. We will release v1.5.3 of TUSZ and v2.0.0 of TUAR in Fall 2021 prior to the symposium.

ACKNOWLEDGMENTS

Research reported in this publication was most recently supported by the National Science Foundation's Industrial Innovation and Partnerships (IIP) Research Experience for Undergraduates award number 1827565. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the official views of any of these organizations.

REFERENCES

[1] 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. *https://doi.org/10.3389/fnins.2016.00196*.

[2] V. Shah et al., "The Temple University Hospital Seizure Detection Corpus," *Frontiers in Neuroinformatics*, vol. 12, pp. 1–6, 2018. *https://doi.org/10.3389/fninf.2018.00083*.

[3] A. Hamid et, al., "The Temple University Artifact Corpus: An Annotated Corpus of EEG Artifacts." in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (SPMB), 2020, pp. 1-3. *https://ieeexplore.ieee.org/document/9353647*.

[4] Y. Roy, R. Iskander, and J. Picone, "The NeurekaTM 2020 Epilepsy Challenge," NeuroTechX, 2020. [Online]. Available: *https://neureka-challenge.com/*. [Accessed: 01-Dec-2021].

[5] S. Rahman, A. Hamid, D. Ochal, I. Obeid, and J. Picone, "Improving the Quality of the TUSZ Corpus," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (SPMB), 2020, pp. 1–5. *https://ieeexplore.ieee.org/document/9353635*.

[6] V. Shah, E. von Weltin, T. Ahsan, I. Obeid, and J. Picone, "On the Use of Non-Experts for Generation of High-Quality Annotations of Seizure Events," Available: *https://www.isip.piconepress.com/publications/unpublished/journals/2019/elsevier_cn/ira.* [Accessed: 01-Dec-2021].

[7] D. Ochal, S. Rahman, S. Ferrell, T. Elseify, I. Obeid, and J. Picone, "The Temple University Hospital EEG Corpus: Annotation Guidelines," Philadelphia, Pennsylvania, USA, 2020. *https://www.isip.piconepress.com/publications/reports/2020/tuh_eeg/annotations/*.

[8] D. Strayhorn, "The Atlas of Adult Electroencephalography," *EEG Atlas Online*, 2014. [Online]. Available: *http://eegatlas-online.com*. [Accessed: 01-Aug-2021].

NEURAL ENGINEERING DATA CONSORTIUM
www.nedcdata.org

College of Engineering
Temple University

Recent Advances in the TUH EEG Corpus: Improving the Interrater Agreement for Artifacts and Epileptiform Events

G. Buckwalter, S. Chhin, S. Rahman, I. Obeid and J. Picone

The Neural Engineering Data Consortium, Temple University

Abstract

- The latest version of TUSZ (v1.5.3) and TUAR (v2.0.0) represents our efforts to review the quality of the annotations based on feedback from the community.
- Enhancements to TUSZ include:
  - High frequency seizures (above 3 Hz frequency) were annotated.
  - The overlap of subjects in training, development, and blind evaluation was removed.
  - All annotations were reviewed by two annotators to ensure there is high interrater agreement among the annotators.
  - Long files were split into smaller sections to eliminate memory issues in the Python code.
- Enhancements to TUAR include:
  - Support of an XML format that allows annotations to be arbitrarily complex and support overlaps in time.
  - Eye movements and other artifacts were annotated with greater precision.
  - Inclusion of artifact bleeding among channels
- The Temple University Hospital Seizure Detection Corpus (TUSZ) and TUH EEG Artifact Corpus (TUAR) are subsets of the TUH EEG Corpus (TUEG).
- TUEG currently includes over 5,000 subscribers.

Introduction

- In TUSZ, every seizure event that occurs in a file has been annotated.
- This latest version of the corpus uses updated annotation standards to reduce the amount of mislabeled background that was contained in seizure files. This helps machine learning algorithms reduce their false alarm rate.
- New documentation for the corpus includes:
  - Annotation guidelines were updated to include high frequency seizures.
  - The overlap of subjects in training, development, and blind evaluation was removed.
  - Files > 1 hour were split into shorter segments to reduce computing memory requirements.
  - Included absence (absz), complex partial (cpsz), generalized (gnsz), focal non-specific (fnsz), simple partial (spsz,) tonic-clonic (tcsz), and tonic (tnsz)
- However, artifacts were not annotated and are a major source of error for machine learning systems.
- In TUAR, every artifact event present in a file has been annotated.
- Our annotation file format was modified to use XML, which enabled multiple artifact tags to be used at one time without causing clutter.

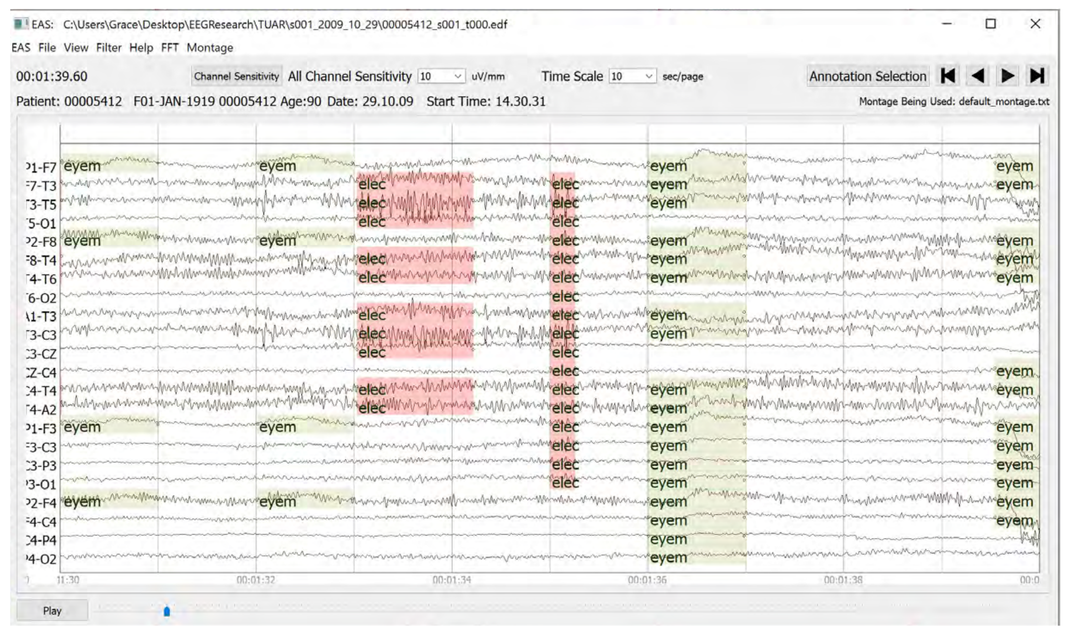
TUH EEG Seizure Corpus (TUSZ: v1.5.3)

- This is a subset of TUH EEG (TUEG) developed to identify seizure activity
- In v1.5.3, there are three subsets; development, training, and blind evaluations
- Statistics for development subset for latest releases:

Development Releases	v1.5.1	v1.5.2	v1.5.3
Patients	50	50	53
Sessions	259	238	341
Files	1152	1012	1832
Seizure Files	309	280	324
Total No. Seizure Events	718	673	1239
Total Duration (Hours)	191	170.3	435.5
Seizure Duration (Hours)	16.9	16.2	21.8

TUH EEG Artifact Corpus (TUAR: v2.1.0)

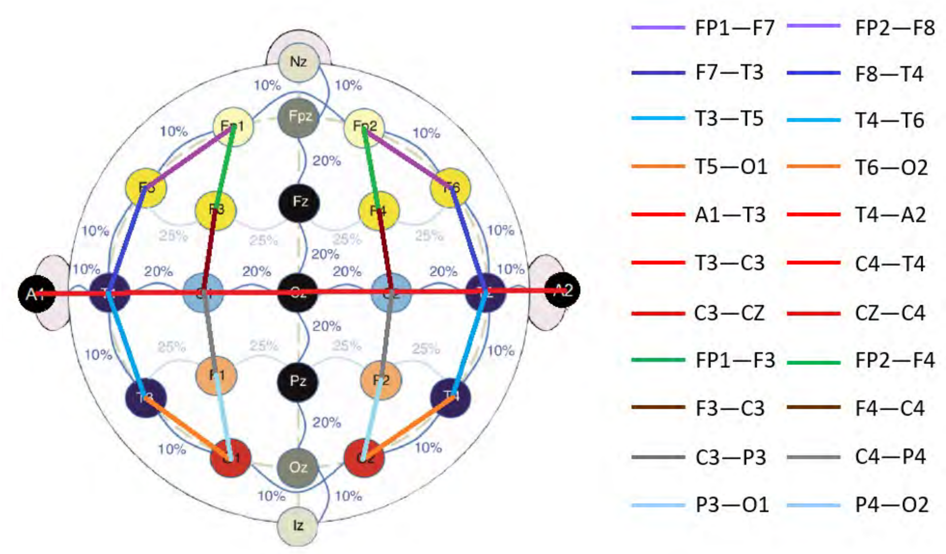
- This is a subset of TUH EEG (TUEG) developed to differentiate artifacts from seizure activity.
- The types of annotation tags include eye movement (eyem), muscle (musc), shivering (shiv), chewing (chew), and electrode (elec).
- In the case of multiple simultaneous artifact events, there are coupled tags (ex: musc+elec).
- Eye movement is often mistaken as seizure activity as seen below:



- TUAR v2.1.0 consists of eye movements that were annotated individually, as opposed to one continuous tag.
- There are 310 files, 259 sessions, and 213 patients in the TUAR Corpus. It has been a popular choice to benchmark the performance of artifact detection.

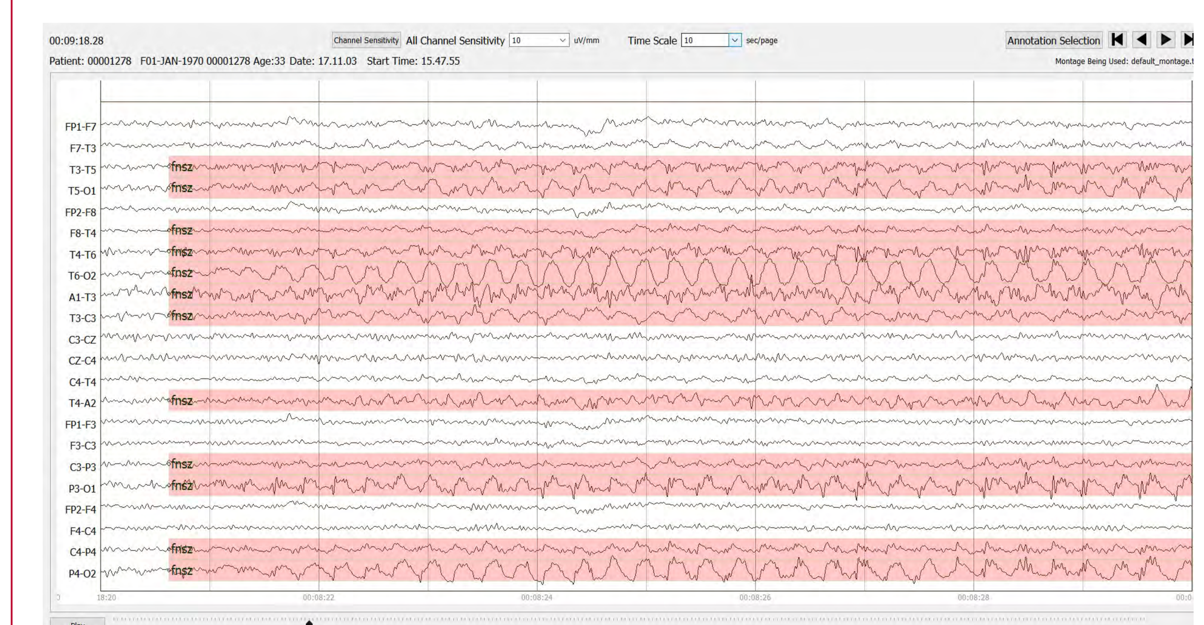
Annotation Process

- In v1.5.2, each file was reviewed by a minimum of three annotators.
- In v1.5.3, the previous annotations from v1.5.2 were reviewed using updated annotation guidelines.
- The files are divided amongst the annotators and are annotated individually.
- After each file has been reviewed once, the annotators review their annotations together. This second step is repeated to ensure complex epileptic form activity present in files is consistently accurate.
- Inter-rater agreement is quite high (kappa > 0.8) for this process.
- Machine learning experiments are underway to assess the impact of the new annotations.



Updated Annotation Guidelines

- The official annotation guidelines used by NEDC can be found here: https://www.isip.piconepress.com/ publications/ reports/2020/tuh_eeg/annotations/.
- High frequency seizures have a frequency greater than 3 Hz. These seizures were not previously annotated in v1.5.2.
- In v1.5.2 there were 718 seizures, while in v1.5.3 there were 843 seizures annotated.
- Instead of using only general seizures (gnsz) and focal seizures (fnsz) tags, the specific types of seizures were also annotated, including absz, cpsz, fnsz, spsz, tcsz, and tnsz
- All annotations were reviewed by two annotators to ensure accuracy and consistency among annotators.



File Formats

- The longest duration files in the corpus can extend for hours, requiring Gbytes of memory in Python, making it difficult to annotate on laptops.
- Long files were split into smaller sections of one hour or less to eliminate these memory issues.
- We also deprecated our rec file format in favor of CSV and XML formats. Our users expressed a strong preference for CSV formats because it is an easy format to manipulate in Python.
- XML files allow more flexible descriptions of the annotations and allow annotations to be represented as directed graphs.
- Below is a typical XML file (00000258_s002_t000):

```
<?xml version="1.0" ?>
<label name="00000258_s002_t000" dtype="parent">
 <endpoints name="endpoints" dtype="list">[[0.0, 1.0255]]</endpoints>
 <probability name="probability" dtype="list">[1.0, 1.0]</probability>
 <tcp_ar name="tcp_ar" dtype="parent">
  <FP1-F7 name="FP1-F7" dtype="*">
   <bckg name="bckg" dtype="parent">
    <endpoints name="endpoints" dtype="list">[[0.0, 1.0255]]</endpoints>
    <probability name="probability" dtype="list">[1.0]</probability>
   </bckg>
  </FP1-F7>
 </tcp_ar>
</label>
```

- The CSV equivalent is:

```
# version = csv_v1.0.0
# montage file: C:\Users\Grace\Desktop\EEGResearch\nedc_annotator_v5.0.3\lib\ned
c_ann_eeg_tools_map_v01.txt
# annotation label file: C:\Users\Grace\Desktop\EEGResearch\nedc_annotator_v5.0.
3\lib\nedc_ann_eeg_tools_map_v01.txt
#
00000258_s002_t000, FP1-F7, 0.0000, 1.0255, bckg, 1.0000
```

Summary

- These corpora and supporting tools are open source and freely available at:
  https://www.isip.piconepress.com/projects/tuh_eeg
  For further information, contact help@nedcdata.org.
- Future release plans include:

Database	Version	Description	Expected Date
TUSZ	v1.5.3	Improved annotations of /dev and elimination of patient overlap	December 2021
TUAR	v2.1.0	Improved annotations of eye movement	December 2021
TUSZ	v1.5.4	Improved annotations of /train and /eval	March 2022
TUEG	v2.0.0	Includes data from 2016-2021	May 2022

Acknowledgements

- Research reported in this publication was most recently supported by the National Science Foundation Partnership for Innovation award no. IIP-1827565. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the official views of any of these organizations.
- We are deeply indebted to Dr. Mercedes Jacobson for her continued support of this project, and to Dr. Vinit Shah, a recent PhD graduate from our program, who was instrumental in training our annotators.