**Leveraging Big Data Resources for Automatic Interpretation of EEGs**

*Christian Ward, Dr. Iyad Obeid, and Dr. Joseph Picone* *Mercedes Jacobson, MD*

The Neural Engineering Data Consortium Department of Neurology

Temple University Temple University School of Medicine

Philadelphia, Pennsylvania 19122 Philadelphia, Pennsylvania 19140

Tel: 215-204-4841 Tel: 215-707-8910

Email: {tuf27176, iobeid, picone}@temple.edu Email: Mercedes.Jacobson@tuhs.temple.edu

The Neural Engineering Data Consortium (NEDC) is being launched to develop big data resources to be used to address the next generation of data-driven research challenges for the neural engineering community. NEDC’s primary mission will be to focus the attention of the research community on a progression of neural engineering research questions and to generate data to support those investigations. A community-wide assessment, funded by a planning grant from the National Science Foundation, is being conducted to define and prioritize the resources required by researchers to fuel innovation. NEDC will broaden participation by making data available to research groups who have significant signal processing expertise but who lack capacity for data generation. This effort is modeled in part after similar successful endeavors, particularly in the human language technology field where the Linguistic Data Consortium (LDC) has led to systematic research and technology advances over a 20‑year span.

NEDC’s first corpus – clinical electroencephalogram (EEG) recordings conducted at Temple University Hospital from 2002 to 2013 – is now under development. The Temple University Hospital EEG Corpus (TUH-EEG) will be the world’s largest publicly available database of clinical EEG data. Although information disclosing a patient’s identity, such as name and corresponding video are being redacted, other demographic information such as gender, age, ethnicity, relevant medical history, and medications will be retained. This corpus will support the development of technology to automatically interpret EEGs in addition to advancing the basic science of what aspects of a patient’s medical record correlate with various pathologies that can be diagnosed from EEG studies. The complete corpus is expected to be freely available in early 2014.

This presentation will consist of three parts: (1) an overview of the planning activities underway to define NEDC’s mission; (2) a description of the TUH EEG Corpus and (3) presentation of some preliminary machine learning experiments on the data.