Hierarchical Event Descriptor (HED) Tags for Analysis of Event-Related EEG Studies

Nima Bigdely-Shamlo[1], Kenneth Kreutz-Delgado[1], Kay Robbins[2], Makoto Miyakoshi[1], Marissa Westerfield[1], Tarik Bel-Bahar[1], Christian Kothe[1], Jessica Hsi[1], Scott Makeig[1]

[1]Swartz Center for Computational Neuroscience, University of California San Diego, La Jolla, CA, U.S.A,
[2]Computer Science Department, University of Texas at San Antonio, San Antonio, TX, U.S.A

***Abstract*— Data from well-designed EEG experiments should find uses beyond initial reports, even when study authors cannot anticipate how it may contribute to future analyses. Several ontologies have been proposed for describing events in cognitive experiments to make data available for re-use and meta-analysis, but none are widely used. One reason for this is that the tools needed to make use of these ontologies are complex, placing a significant burden on experimenters while not providing any immediate reward for their efforts. Here we propose an extensible, user-friendly experiment event tagging method built on the BrainMap and CogPO ontologies and similar to the object tagging style used extensively on the Web. Hierarchical Event Descriptor (HED) tags, a hierarchy of standard and extended descriptors for EEG experimental events, provide a uniform human- and machine-readable interface facilitating use of an underlying event-description ontology during EEG data acquisition, analysis, and sharing. HED tags may be used to mark and annotate all known events in an experimental session. We describe an available real-time EEG experiment control and recording system that uses HED tags for annotation, transmission and storage of detailed information about events in EEG experiments.**

I. Introduction

The current period in the history of science has been called the era of Big Data collection and analysis. Annotated data can be highly useful to meta-analysis. Hence there have been several efforts to standardize terminologies used to describe cognitive paradigms by developing formal database ontologies (e.g., Nemo [1], BrainMap [2], CogPO [3], and NeroLex [4]), but currently none of these is often used to describe EEG events in publications. We believe key hindrances are:

(1) *Complexity of appearance and difficulty of use:* Although formal ontologies developed in the OWL format [5] are elegant and can be readily processed by computers, their apparent complexity discourages human casual use. Tools to interact with and make use of such ontologies are complicated, and using them may require learning a large number of detailed concepts, standards, and file formats.

(2) *Lack of immediate reward for use:* Most neurophysiological database tools do not provide clear and immediate benefits to researchers who perform the work required to donate their data to the resource – i.e., they give little or no tangible reward to make researchers feel it worthwhile to undertake the work of annotating and uploading data for (their own and/or others') further use.

II. Hierarchical Event Descriptors

To address problem (1) above, we have adopted a popular object tagging style used extensively on the Web (for example, for image tags on Flicker.com and video tags on YouTube.com). *Hierarchical Event Descriptor (HED*) tags are a hierarchy of standard and extended descriptors for EEG experiment events. The HED system includes a base set of hierarchically organized descriptor tags, in part adapted from the BrainMap and CogPO ontologies [2, 3], that can be used to describe many types of EEG experiment events in a uniform (though easily extensible) human- and machine-readable manner. The main contribution of the HED tagging system is to offer a user-friendly interface for use of the underlying event description ontology in EEG acquisition and analysis workflows.

Another goal of HED tagging is to support EEG data analysis and meta-analysis by enabling automated discovery of appropriate statistical designs in complex EEG studies including many types of known events. Using HED tags could provide an immediate reward to researchers by simplifying and automating their analysis workflow, thus addressing (2) above.

The hierarchical structure of the HED tags makes it easy to search through variations of the same type of event across studies (enabling EEG data meta-analyses), while preserving the unique details of each event type. For example, an event marking the presentation of a visual feedback stimulus in EEG Study A may present a red circle to the participant on a black screen background, while in Study B the visual feedback stimulus is a blue rectangle on a white screen background. In HED syntax, these event types can be described by two HED strings (collections of comma-separated HED tags) as follows:

In Study A:

```
Stimulus/Feedback,
Stimulus/Visual/Uniform Color/Red,
Stimulus/Visual/Shape/Ellipse/Circle/Height/2-
deg,
```

```
Stimulus/Visual/Shape/Ellipse/Circle/Width/2-deg,
Stimulus/Visual/Background/Uniform Color/Black
```

In Study B:

```
Stimulus/Feedback,
Stimulus/Visual/ Uniform Color/Blue,
Stimulus/Visual/Shape/Rectangle/Height/2-deg,
Stimulus/Visual/Shape/Rectangle/Width/3-deg,
Stimulus/Visual/Background/Uniform Color/White
```

These descriptors explicitly capture both salient commonalities across and differences between the two event types. If the feedback events were accompanied by an auditory beep (500 Hz, 25-dB), the following tags might be added:

```
Stimulus/Auditory/Loudness/25-dB,
Stimulus/Auditory/Tone/500-Hz,
Stimulus/Auditory/Tone/Ramp Up/10-ms,
Stimulus/Auditory/Tone/Ramp Down/10-ms
```

While higher levels of the HED hierarchy are intended to be fixed (i.e., revised infrequently with discrete versioning based on community feedback), lower levels may be extended without restriction to describe any event type to any desired level of detail. For example, the tag *Stimulus/Visual/Shape/Ellipse/Circle* may be extended by adding */Filled* at the end to provide more information about the circle.

In addition to describing delivered stimuli, HED tags can also describe subject actions (e.g., button presses, swipes, saccades, etc.), subject and task states (e.g., drowsy, attend visual, etc.), and combinations of these (e.g., feedback tones produced in immediate response to button presses by a drowsy subject). Since HED tags can be easily interpreted by computer applications, HED tagging can facilitate search and inference of event-related EEG dynamics across multiple studies, when and where available. Also, HED tags organize events from a study into a logical hierarchy so they can be more easily analyzed. For example, several event subtypes may be aggregated into a more general event type that can then be compared to other event types.

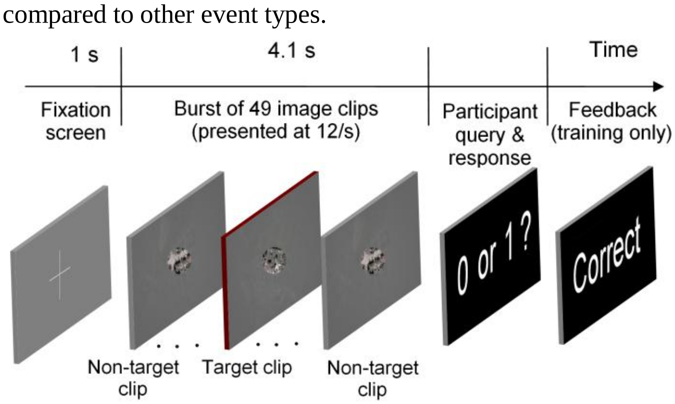

Figure. 1 Trial schematic for an RSVP experiment [6].

A. *Example Tagging for RSVP*

Here we explain the use of HED tags to describe events in a sample Rapid Serial Visual Presentation (RSVP) study of target recognition in satellite imagery [6] using HED 1.31 specification available at [7]. Each recording session of this experiment comprised of 504 4.9-s image bursts of 49 oval image clips from a large satellite image of London presented at a rate of 12/s. Some (60%) of these bursts contained one image in which a target white airplane shape was introduced at a random position and orientation. Following each burst, subjects were asked to press one of two buttons to indicate whether or not they had detected a target airplane in any burst image clip. Figure 1 shows a time line of each RSVP burst. For further details see [6].

Now we use HED tags to form HED strings describing events (from left to right) in Figure 1. Display of a silver fixation cross on a gray background:

```
Stimulus/Visual/Shape/Cross,
Stimulus/Visual/Uniform Color/Silver,
Stimulus/Visual/Achromatic,
Stimulus/Visual/Screen Location/Center,
Stimulus/Visual/Background/Uniform Color/Gray,
Stimulus/Visual/Fixation Point,
Stimulus/Instruction/Fixate
```

Non-Target image presentation event:

```
Stimulus/Visual/Achromatic,
Stimulus/Expected/Non-Target
```

Target image presentation event:

```
Stimulus/Visual/Achromatic,
Stimulus/Target
```

Presentation of visual cue asking the participant whether (s)he has detected a target airplane image ('0 or 1?'):

```
Stimulus/Visual/Language,
Stimulus/Instruction/Count,
Stimulus/Visual/Uniform Color/White,
Stimulus/Visual/Achromatic,
Stimulus/Visual/Background/Uniform Color/Black
```

Participant answers the question by pressing one of two buttons (here, to indicate a response of '1') with his/her right hand:

```
Response/Button Press,
Response/Hand/Right Hand/Index Finger,
Response/Count/1
```

Presentation of visual feedback ('Correct' or 'Incorrect') cue in training trials:

```
Stimulus/Visual/Language/Word/Noun,
Stimulus/Visual/Language/Latin/English,
Stimulus/Feedback/Correct, (or /Incorrect)
Stimulus/Visual/Uniform Color/White,
Stimulus/Visual/Achromatic,
Stimulus/Visual/Background/Uniform Color/Black
```

Finally, we specify the paradigm as:

```
Paradigm/Rapid Serial Visual Presentation/Visual Target Detection Paradigm,

Paradigm/Oddball discrimination paradigm/Visual oddball paradigm
```

B. HED Grammar

HED tags consist of a series of identifiers separated by the forward slash '/' character. A HED identifier may contain any characters except the characters '/' (forward slash), ',' (comma), ';' (semicolon), and '"' (quotes). To use these reserved characters in a HED identifier, the HED identifier may be wrapped in a pair of double-quote (") characters. To use the quote character inside a quoted HED, use '\"' (backslash double-quote).

Sometimes an event may be associated with more than one stimulus or response – for example when a red circle is presented on screen left and, at the same moment a blue rectangle is displayed on screen right. In such cases, parentheses can be used to group together HED tags associated with each stimulus. In this example:

```
(Stimulus/Visual/Shape/Ellipse/Circle,
Stimulus/Visual/Uniform Color/Red,
Stimulus/Visual/Screen Location/Left),

(Stimulus/Visual/Shape/Ellipse/Rectangle,
Stimulus/Visual/Uniform Color/Blue,
Stimulus/Visual/Screen Location/Right)
```

III. HED Tools and Resources

MATLAB (The Mathworks, Inc.) tools for basic HED tag operations, such as validation, auto-completion and comparisons between pairs of HED tags and strings have been developed and are available at [7].

A. *Community HED Tagging*

Community tagging tools (described in a parallel submission [8]) have been developed that allow users to tag events based on their types and other attributes. These tools allow users to create HED tag overlays that can be easily edited to allow retagging. The tools include a HED database with usage counts, so that a group of users can collaborate to develop new tags. These tools can be called from the MATLAB command line or from the EEGLAB [9] menu.

B. *Real-time HED Tagging of Experimental Conditions*

Most EEG acquisition hardware only provides a single time series channel for recording the (integer-coded) type of experimental event, associated with the concurrent EEG data frame. This limits the number of distinct events that can be recorded and makes it necessary to associate the data with a table that maps these 'event type' indices to their human-readable descriptions.

However, a current trend in cognitive neuroscience is towards performing more naturalistic and less constrained experiments, for example experiments in which participants play games in ambulatory settings. Because of the multiplicative interactions of behavioral and contextual dimensions of interest in these experiments, the number of potential event types can be quite large and can hardly be captured by a one-dimensional (often 1- or 2-byte) event number channel. Even when it is possible to sacrifice encoding experiment details and only use a single channel for recording this information, the resulting mapping table could become quite large and difficult to maintain and use for statistical analyses.

An alternative approach is for the experiment control and data recording software to use HED tags to fully encode all aspects of interest for each instance of an experimental event, directly sending the resulting HED strings, in real time, to the data acquisition system. These strings will then be recorded synchronous to the recorded EEG data.

We (C. Kothe *et al*.) have developed, tested, and now use in practice a real-time interactive experiment control and data recording system that implements this approach using HED strings for event description. The system consists of the Simulation and Neuroscience Application Platform (SNAP) for real-time experimental control [10], the lab streaming layer (LSL) framework for synchronous multimodal data transfer [11], and the Extensible Data Format (XDF) file format for data storage [12].

SNAP is a python-based experiment control framework that can send HED strings to LSL, a (C++-coded) real-time data collection and distribution system, to be recorded with acquired EEG data (in our laboratory from standard Biosemi EEG hardware) in an XDF file format capable of synchronously recording multi-channel, multi-stream data that is heterogeneous in both type and sampling rate. Figure 2 shows the links between components of this system.

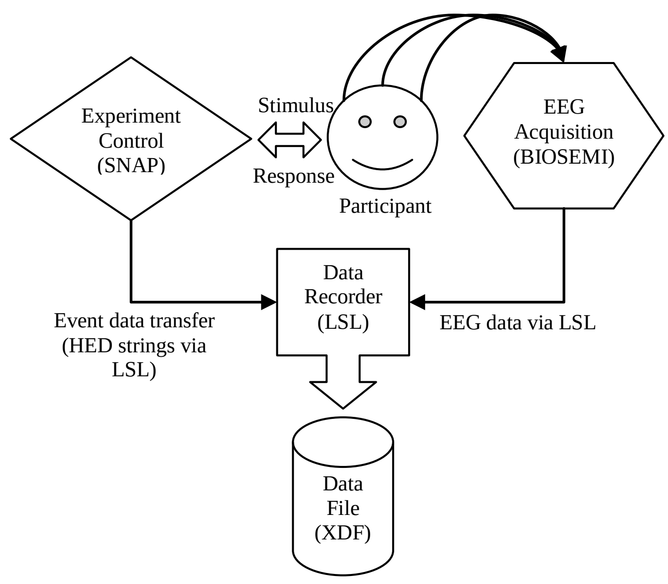

Figure 2. A system for real-time HED tagging and synchronous recording of EEG data and events. LSL tools including drivers for many types of input devices are available at [11].

C. *Extended Use of HED Tags for Study Meta-Data*

We have also developed a companion XML-based specification, called the *EEG Study Schema* (ESS) to hold all the information necessary to analyze an EEG study, e.g. subject gender, handedness, age and group associations, task and paradigm description, etc. in a format that is both machine and human readable. To achieve this goal, ESS relies on HED descriptions of experimental events and tasks that are embedded in an ESS XML document. When ESS XML files are viewed in current web browsers (e.g., FireFox) they are automatically formatted as readable reports using a provided XSLT style sheet. For more information about ESS see [13].

Using the ESS/HED system we have documented 18 laboratory studies comprising 388 data recording sessions. Five of these studies are publicly available at our online EEG study repository HeadIT [14]. Figure 3 shows a hierarchical representation of HED-tagged event types from all 18 studies (most event types typically being in common to all study sessions). Values in parentheses show the number of unique event types from all studies that match the HED tag associated with each level of the hierarchy.

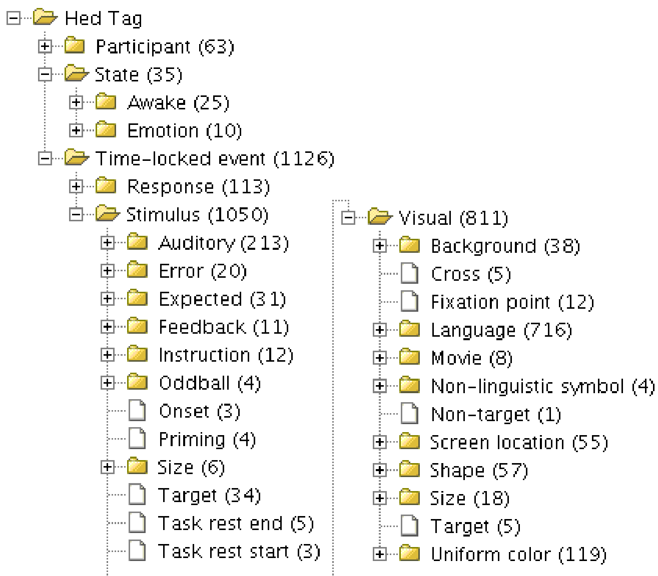

Figure 3. Hierarchical representation of HED-tagged event types from 18 archived studies. Values in parentheses show the number of unique event types, across all studies, that match the HED tag associated with each level of the hierarchy.

IV. Discussion

A future direction is to make formal connections between HED tags and ontological terms defined in Neurolex [4]. Another extension is to classify the type of relationship between HED child nodes and their parents ('is-a' for event subtypes such as `Ellipse/Circle`, versus 'has-a' for properties like `Shape/Color`). This should enable more complete automated statistical analysis of brain dynamics associated with HED-specified cognitive events and states. Finally, tools that immediately reward researchers for annotating and uploading their data, by returning useful information to them about it, may generate the researcher interest needed to amass a sufficient quantity of data to allow large scale EEG data mining for a range of purposes.

Acknowledgement

This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0022, by NIH grant 1R01MH084819-03, and by a gift from The Swartz Foundation (Old Field NY). The views and the conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S Government. The U.S Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] *Neural Electro Magnetic Ontologies (NEMO)*. Available: http://nemo.nic.uoregon.edu/
[2] P. T. Fox, A. R. Laird, S. P. Fox, P. M. Fox, A. M. Uecker, M. Crank, S. F. Koenig, and J. L. Lancaster, "BrainMap taxonomy of experimental design: description and evaluation," *Human brain mapping*, vol. 25, pp. 185-98, May 2005.
[3] *Cognitive Paradigm Ontology (CogPO)*. Available: http://wiki.cogpo.org/index.php?title=Main_Page
[4] *The Neuroscience Lexicon*. Available: http://neurolex.org/wiki/Behavioral_Paradigm_Hierarchy
[5] *OWL 2 Web Ontology Language Document Overview*. Available: http://www.w3.org/TR/owl2-overview/
[6] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, "Brain Activity-Based Image Classification From Rapid Serial Visual Presentation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, pp. 432-441, Oct 2008.
[7] *HeadIT tools*. Available: http://sccn.ucsd.edu/wiki/HED
[8] Rognon, R. Strautman, J. Lauren, N. Bigdely-Shamlo, S. Makeig, T. Johnson, and K. Robbins, "CTAGGER: Semi-structured community tagging for annotation and data-mining in event-rich contexts " submitted to the IEEE GlobalSIP, Austin, TX, 2013.
[9] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, pp. 9-21, Mar 15 2004.
[10] *Simulation and Neuroscience Application Platform (SNAP)*. Available: https://github.com/sccn/SNAP
[11] *Lab Streaming Layer (LSL)*. Available: https://code.google.com/p/labstreaminglayer/
[12] *Extensible Data Format (XDF)*. Available: https://code.google.com/p/xdf/
[13] *EEG Study Schema (ESS)*. Available: http://sccn.ucsd.edu/wiki/ESS
[14] *Human Electrophysiology, Anatomic Data, and Integrated Tools (HeadIT)*. Available: http://headit.org/