USDA / Auburn CS+Math Clinic Minutes

**Date and Time**: 30 August 2024, 4:00 pm

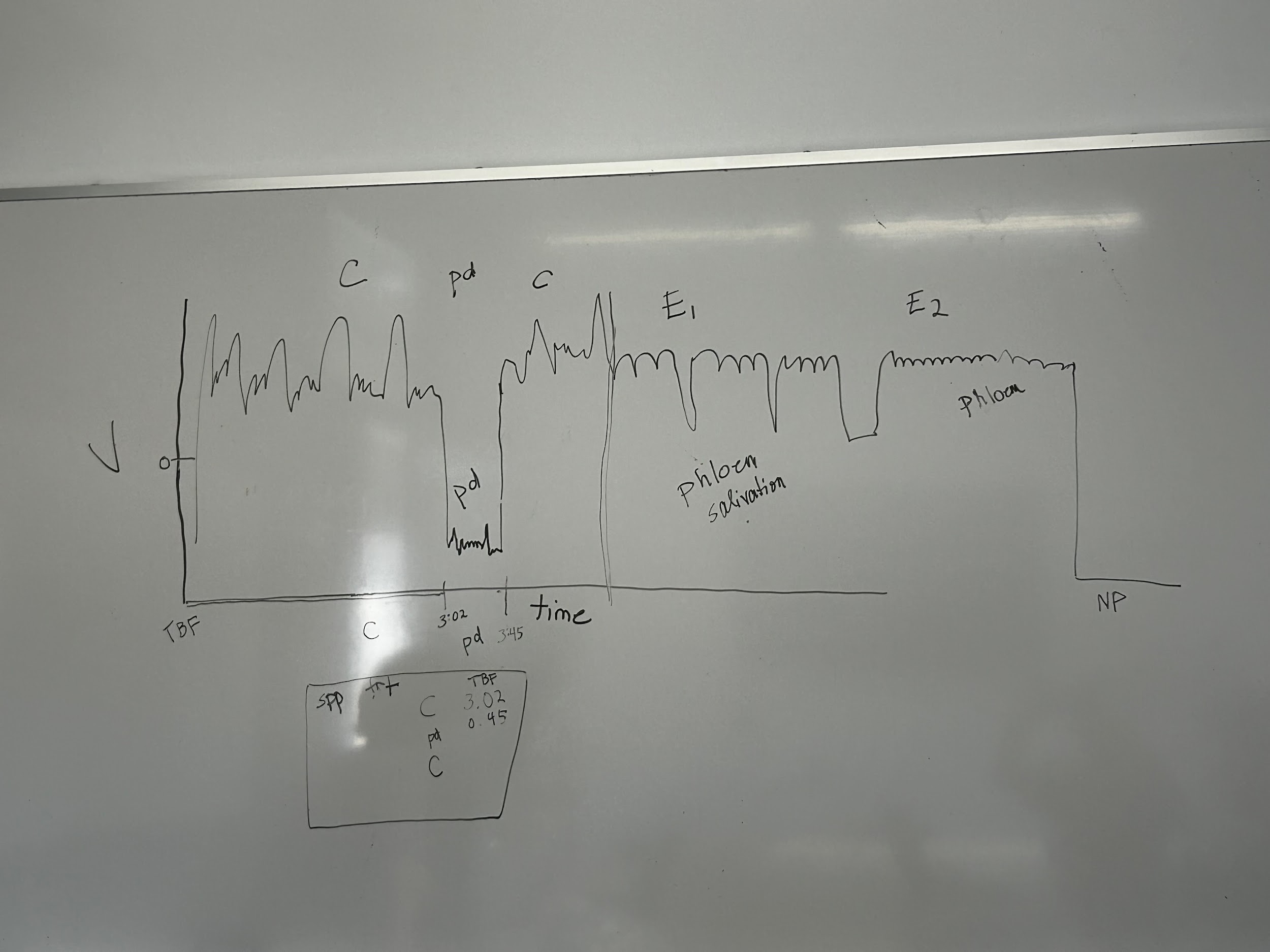
**Location**: Shanahan 3466, Harvey Mudd College

**Attendees**: Dr. Backus, Dr. Reif (via phone), Prof. Hope, Zachary, Milo, Lillian, Mehrezat, Devanshi

**Objectives**

* Discuss project vision, requirements, goals
* Discuss which logo should be used

**Minutes** (taken by Zachary)

* Introductions (4:00)
* Discussion of vision (4:20)
  + Vocabulary note
    - Measurement: what computer scientists would call an “annotation” or “label”
  + EPG Data: Voltage vs Time
    - Whiteboard Picture: 
    - Dr. Backus drew this to illustrate what an EPG waveform looks like and how the current labeling system works. This particular example is representative of aphid data (plot) and annotations (table under plot)
  + Dr. Backus currently uses Windaq to measure (annotate) plots like the above by manually selecting intervals and assigning a label. This is incredibly time consuming.
  + This clinic team is **not** to work on improving the interface with the EPG system, it is only to create an automated labeling tool. The interface of the EPG system is the responsibility of the CS students on the engineering clinic team.
  + Dr. Backus is working on getting us aphid data, she was not able to get it to us today due to her laptop being dead. We will need to work on getting the data from her.
  + We need a program that is flexible enough to work across any species of bug that can produce meaningful EPG data.
    - We should get an example of what this output looks like so we can match it.
  + Four or five labs have attempted this, but the only one that seems to really have been successful is a Chinese former post-doc of Dr. Backus’. Americans are not currently allowed to collaborate with Chinese scientists so we do not have access to this method.
    - There is a paper describing the system but only in general terms.
  + Variation: We should expect to see person-person and device-device variation. Aphids are very well studied though so their dataset should be the best. Different insects will have different numbers of labels.
  + This project is funded by National Bio and Agro Defense Facility (NBAF). Incredibly high biological security. Only facility in the United States that can work with the worst livestock diseases in the world in the United States. Their mission is with mosquitos so that is our eventual mission.
  + The way Dr. Backus sees this working is that we have a program that can take in training data, a dataset we want labeled, and it should produce labels for the waveform.
    - Input: voltage over time data
    - Output: two column dataframe with columns [label, start time]
* Dr. Reif joins via speaker phone (5:10)
  + Dr. Backus and Zach gave a summary of the meeting so far to Dr. Reif.
  + Dr. Reif emphasized that the priority is the mosquito dataset. It could be a misstep to spend time on the aphid dataset. The funder’s preference is likely to focus entirely on mosquitos. There is also funding for ticks so once we get those recordings we should work on them (this will not likely happen for a few months).
  + They will get the data to us sometime soon (ideally September 9 or 10).
  + Our liaison meeting will be 8:30 Thursday mornings
* Dr. Backus and Dr. Reif leave (5:40)
* Clinic team and Prof. Hope tentatively decided on a weekly meeting from 4:15-5:15 on Mondays in Prof. Hope’s office

**Takeaways**

* The clinic team’s main goal is to produce a tool that can take in voltage over time data and automatically label the data by behavior, particularly with respect to mosquitos.

**Action Items**

* Mehrezat will send out an email with a calendar invite to liaisons.
* Mehrezat will send out a google calendar invite for weekly meetings.
* Liaison email alias needs to be updated to include all liaisons.
* Liaisons will give the clinic team a labeled EPG dataset ideally by September 9 or 10.
* Liaisons will give the clinic team papers and software related to auto-labeling from Dr. Backus.
* Upon receipt of an EPG dataset in the .windaq format, the clinic team will write a macro script to convert all of them to something usable.
* Liaisons will give team acknowledgement numbers for NABF
* Dr. Backus will send the clinic team the sponsor’s logo(s) they should be using on records and presentations.