

901502 - Passive Acoustic Sensing

Week 51

1. Danelle Cline
 - a. Team meeting
2. Carlos Rueda
 - a. Group meeting

Week 50

1. Carlos Rueda
 - a. ASA [Poster](#) presented today 12/9. Chat session with good comments/ideas, and a few requests for "links to the poster."

Week 49

1. Carlos Rueda
 - a. Finalizing ASA poster. Many thanks John for the productive discussions.

Week 48

1. Brent Jones
 - a. Processed hydrophone click data for May2020 - Present

Week 47

1. Carlos Rueda
 - a. More experiments, ASA poster preparation, and advancing technical report.

Week 46

1. Danelle Cline
 - a. Finished metadata and upload scripts for AWS Open Data Sharing program initial 2.5 month data set. Upload of raw data underway. So far everything has worked well. Prototyped web design for communicating with ITD external web changes to serve data tutorials. John and I are working on prototyping notebooks in Colab and EC2 Jupyter stacks to demonstrate how to use the data.
 - b. Humpbacks have been busy singing in the evenings. Check-out the [live stream](#) if you need some relief from the chaos.
 - c. Call planned next week with folks at the Earth Species Project to discuss our plans for data sharing.

Week 45

1. Carlos Rueda
 - a. More literature review, classification and comparisons with a nice external [tool](#) I just recently found. Some ASA poster preparation.

Week 44

1. Danelle Cline

- a. Working on scripts to describe and share our entire 5-year archive in the Amazon Open Data Sharing program. Cloud push and testing planned this week for 2020 data.
 - b. Found a normalization issue in the blue whale call prediction code which reduced the effective prediction accuracy. Reprocessing the blue A and D calls with this patch. Accuracy is now 93 and 95% for each model respectively. Models take approximately 30 minutes to train.
2. Carlos Rueda
 - a. More HMM, Naive Bayes, and Markov classification exercises. Generating and reviewing scatter plots of classification rank vs. sequence length. Reviewing literature, resuming writing of technical report, and preparing ASA poster.

Week 43

1. Kevin Gomes
 - a. Some AWS support for Cal Poly students
2. Carlos Rueda
 - a. Preparing for a next round of classification exercises and also for the ASA poster.

Week 42

1. Kevin Gomes
 - a. Helping with AWS related stuff
2. Brent Jones
 - a. Uploading odontecete click data to AWS
3. Carlos Rueda
 - a. Met with John to catch up with status and start preparing ASA poster.

Week 41

1. Danelle Cline
 - a. Testing ported tensorflow to pytorch blue whale classification code
 - b. Assisting with versioning and sharing click data with Cal Poly AI class
 - c. Discussion and planning with John Ryan and Kevin Gomes on how to best capture metadata and share with the AWS Open Data Sharing program
 - d. Working on cleaning up notebooks to share with POGO workshop participants & AWS Open Data Sharing program
2. Kevin Gomes
 - a. Working on data sharing

Week 40

1. Kevin Gomes
 - a. Moved LTSA web site to S3 and am evaluating performance and looking to enable SSL, Route53 DNS entry and enabling CloudFront, then testing performance again.
2. Brent Jones
 - a. Setting up collaboration with CalPoly students who will be working with hydrophone data for their class project.

Week 39

1. Danelle Cline
 - a. AWS upload throttle test
 - b. Discussion with Cal Poly about potential student projects
 - c. Team meeting

2. Kevin Gomes
 - a. Some meetings and small amount of AWS stuff
3. Brent Jones
 - a. Converting high pass filter in Julia to python. Investigating results
4. Carlos Rueda
 - a. Group meeting

Week 38

1. Kevin Gomes
 - a. Meeting to discuss AWS plans

Week 37

1. Danelle Cline
 - a. Restarted shoutcast stream which stopped after running fairly robustly for some time. There were no errors in the log file so it remains a mystery.
 - b. Fixed email notification on recording machine which was failing due to secondary authentication checks.

Week 36

1. Danelle Cline
 - a. Troubleshooting issues with streaming audio failure.
 - b. Trying to keep up with the AWS Data Sharing chatter
 - c. Blue whale D call predictions ready to process. Just need to free up GPU resources to start.
 - d. Discussion with Daniel DeLeon about inference tests results on Aria 10 FPGA and ideas for his masters studies.
2. Kevin Gomes
 - a. Created an AWS account for the public data sharing, registered it with Amazon's Organization, created user and set up security policies to limit to S3 activities only.
 - b. Created S3 buckets using CloudFormation templates provided by the AWS data sharing team

Week 35

1. Danelle Cline
 - a. AWS data sharing program correspondence. Many thanks to Kevin Gomes and the IS department for their help as we forge into the great Cloud.
 - b. Added coherence and VGGish embeddings plots of humpback topic model output. The dispersion of the VGGish embedding looks like a better discriminator than coherence for model optimization.
 - c. Blue whale A call predictions across the entire 5-year time-series produced >200K calls. Next week blue processing jobs for whale D call predictions will be put in the queue.
2. Kevin Gomes
 - a. Created an AWS account and AWS connected the account to their data sharing organization where they will pay to host the Soundscape data for public consumption.
3. Brent Jones
 - a. Running ML sperm detector on all clicks for a given month. Continuing to tune algorithm
 - b. Spun up virtual Linux machine for parallel processing hydrophone data

Week 34

1. Carlos Rueda
 - a. Meeting. Resuming work on technical report.

Week 33

1. Carlos Rueda
 - a. Submitted abstract to ASA (thanks John!).
 - b. Putting together a technical report about the various VQ/HMM experimentation exercises.

Week 32

1. Danelle Cline
 - a. Fixing an issue with dockerized deployment of blue whale model in preparation for deployment on 5yr time-series with VGGskeleton model. Processed a few days for collaborator, Will Oestreich to review; he reports 100% accuracy on test files. Model was around 92% on the validation set.
 - b. Summer intern discussions and drafting up collaboration agreement.
 - c. Ongoing decimation filtering discussions with the team.
2. Carlos Rueda
 - a. Reviewing some LPC theory and adding plots based on LPC spectral envelope along with regular the spectrograms in classification results. Planning to run exercises with LPC prediction order 20 according to typical rule of thumb based on sampling frequency.

Week 31

1. Danelle Cline
 - a. Tested optimized decimation filter with shorter filters and disseminated to team for inspection. New filters are ~20% faster in the most rate-limiting step for the soundscape analysis. We think the faster decimation is working but needs a time adjustment for the filter lengths which P. McGill will investigate.
 - b. Meetings with summer intern, Madison. She's doing a great job exploring topic coherence methods. Added similarity test for her to find two runs with similar topics based on [ARI](#) for her coherence analysis.
 - c. Doing some clean-up of blue whale batch inference code.

Week 28

1. Danelle Cline
 - a. Conference with Carrie Wall, an oceanographer who leads the fisheries acoustic archives at the NOAA National Centers for Environmental Information (NCEI). Discussed our blue whale research and methods for automated analysis. Data is in hand and experiments are planned to test with our signal processing and machine learning methods.
 - b. Testing single day case study John identified as a high priority for decimation testing. Day has time drifts and multiple irregular drop-outs. Getting close to a robust solution to handle the messy data.
 - c. Started on testing Yanwu's optimized filter parameters in our decimation notebook. These should result in a speed-up in our current rate limiting step of decimating the high-frequency data.
 - d. Topic modeling sweeps for summer intern and ongoing discussions 3x per week about progress on coherence results. Many thanks to Ben Raanan and Carlos Rueda for helping when needed with this effort.
 - e. Project proposal input
2. Carlos Rueda
 - a. Meetings

Week 26

1. Danelle Cline
 - a. Humpback whale song topic model sweeps for Dirichlet hyperparameters: [gamma](#), [alpha/beta](#). Code for that is here: <https://bitbucket.org/mbari/soundscape-humpback-topicmodel>. Expect many more

model sweeps in the coming weeks to fine-tune the final parameters. Looking at coherence in addition to perplexity as a measurement of quality.

- b. Summer intern discussions on coherence, PCEN preprocessing etc.
- c. Staged 5-year decimated data in AWS to share with collaborator Will Oestreich on Blue whale A/D call analysis. It was surprisingly fast to upload the entire 2kHz dataset.
- d. Experimenting with prisma/graphql database with sound metadata.

- 2. Carlos Rueda
 - a. Meeting

Week 25

- 1. Carlos Rueda
 - a. Meetings

Week 24

- 1. Danelle Cline
 - a. Fixed an issue with wav cache read collision in decimation job
 - b. Created postgres development database for MARS wav file metadata using prisma/graphql. Really liking the schema evolution and tools in the software stack.
 - c. Added coherence image enhancement to Blue whale classifier and regenerating data for training.
 - d. Reviving topic model code in preparation for student project efforts.
 - e. Student discussions
- 2. Brent Jones
 - a. Meetings

Week 23

- 1. Danelle Cline
 - a. Code optimization of [blue whale code](#)
 - b. Found issue with missing data in decimation files. Resubmitting backlog jobs to recompute those.
 - c. McGill and Zhang are working on further optimizations to the decimation filters. This should be useful for various workflows that need different levels of decimated data.
- 2. Brent Jones
 - a. Cleaning up documentation in prep for intern

Week 22

- 1. Brent Jones
 - a. Beginning to deal with MARS echosounder clicks that started in 2019. Investigating traditional signal processing tools and a ML detector
- 2. Carlos Rueda
 - a. More experimentation with classification. Wrote python scripts to help [visualize the results](#) along with the original wav data (spectrogram).

Week 21

- 1. Danelle Cline
 - a. Centos8 upgrade
 - b. Finished decimating 5-year time-series and now investigating some 10-minute gaps in decimation likely due to time drift.
- 2. Brent Jones
 - a. Tuning Sperm click detector

- b. Working on a preliminary python processing pipeline for daily sperm click analysis
 - c. Prepping for intern
- 3. Carlos Rueda
 - a. With John, performing labelling analysis based on HMM classification results.
 - b. ECOZ2 executable binaries for Linux and MacOS now automatically [built and released](#) on Github. (See overhead item).

Week 20

- 1. Danelle Cline
 - a. CentOS8 migration of HTCondor pool master machine condormastervm to htcondormaster
 - b. Teleconference with AWS team about open data sharing program
 - c. Weekly team meeting and monitoring of decimation jobs
- 2. Brent Jones
 - a. Reprocessing data using inter-click interval. Found more instances of possible sperm clicks
 - b. Documenting software in prep for summer intern
- 3. Carlos Rueda
 - a. More parameter tuning in [HMM training](#). Interesting new insights in that some parameters can actually be given smaller values to improve performance while avoiding overfitting.

Week 19

- 1. Danelle Cline
 - a. AWS data sharing teleconference to discuss rollout of 2020 data with tutorials
 - b. Computing 2kHz decimated time series with UTC daily files
 - c. Working on HTCondor pool CentOS8 upgrade
- 2. Brent Jones
 - a. Investigating ways to properly record inter-click interval into our analysis. Hopefully dont have to redo complete processing
 - b. Helping forecast size req's as hydrophone data accrues.

Week 18

- 1. Carlos Rueda
 - a. Meeting
 - b. Some catching up regarding improved compute performance, see overhead/personal bullet.

Week 17

- 1. Danelle Cline
 - a. Found and fixed an issue in decimation for edge case with multiple small discontinuous files
 - b. Reprocessing 2020 data with UTC aligned day in preparation for AWS Data Sharing and pre/post COVID soundscape analysis.
- 2. Brent Jones
 - a. Continuing to go through monthly data looking for sperm clicks in an effort to build a ML training set
 - b. Begin LTSA monthly processing for MAR20 and APR20. Want to see how background noise has changed since the covid lockdown
- 3. Carlos Rueda
 - a. Speeding up some of the whale song unit analyses via [multi-threading](#).

Week 16

- 1. Brent Jones

- a. Training a CNN to hopefully detect Sperm Whale clicks. Finished processing 5 years of hydrophone data. Documenting process
2. Carlos Rueda
 - a. Group meeting. Presented recent [HMM exercise](#) on 4.5hr whale song unit data.

Week 15

1. Brent Jones
 - a. Still babysitting reprocessing of all hydrophone data for clicks. Had IS setup a gigantic VM. Installed the software and sicked it on the data. Seems to have memory issues a few times a day that requires restarting processing
 - b. Isolated a few sperm whale clicks from the data. Figuring out a way to mine the data for more so I can create a training set for a CNN
2. Carlos Rueda
 - a. Advanced [cffi](#)-based [python wrapper](#) (with a [callback added to C code](#)) and with that, ran an [initial HMM training exercise with wandb](#) involving a metric (sum of log of probabilities).

Week 14

1. Brent Jones
 - a. Discovered current hydrophoneclick detector pre-filters out Sperm whale clicks (I think, still need to verify). Now having fun reprocessing a metric $\#\$%$ load of hydrophone data.
2. Carlos Rueda
 - a. Ran HMM training/evaluation sessions with several [parameter variations](#).
 - b. Similarly as with a set of whale song unit files and labelling of a few months ago, I also ran a [clustering-based classification](#) on the more recent 4.5 recording and John's labelling.
 - c. Exploring possible tools/services to facilitate model tuning, including [wandb](#) (thanks Danelle for the references), [comet.ml](#) / [mlflow](#), [dvc](#) (this one promising but not yet as useful after some investigation, including an actual exercise and putting together a [repo](#) at dagshub.com).

Moving forward, writing a [python wrapper](#) for the software to facilitate use of those various tools.

Week 13

1. Danelle Cline
 - a. Restarted rsync and Shoutcast audio stream after reboot following IS patch failures. Added reboot check to crontab.
2. Brent Jones
 - a. Processing hydrophone data from Oct 2016 where we have known Sperm whale clicks. Hoping to detect clicks with ML algorithms

Week 12

1. Kevin Gomes
 - a. Short debugging session when audio stream went down
2. Carlos Rueda
 - a. Along with various code adjustments (incl improved command-line interface using Rust), completed a first pass of HMM based classification on the 4.5-hour whale song unit recording provided and labelled by John: <https://github.com/ecoz2/ecoz2-whale/tree/master/exerc01>. The accuracy on test data (not training data) is now 75.7%, a significant increase wrt to a previous exercise (57.3%) involving a different set of recordings and [preliminary labelling](#).

Week 11

1. Carlos Rueda
 - a. Regular meeting (Danelle's presentation on PCEN).
 - b. Applying VQ and HMM on labelled data on the 4.5hr recording file provided by John.

Week 10

1. Danelle Cline
 - a. Working on streamable PCEN implementation notebook
 - b. Retraining blue whale call classifiers with improved PCEN parameters
 - c. Reviewing summer intern applications

Week 9

1. Danelle Cline
 - a. Instrumented [topic model](#) with wandb logging, ran two sweeps for K-means over different PCEN parameters, [first](#) broad then [fine](#). Next steps are to run full model sweeps with best PCEN parameters.
2. Carlos Rueda
 - a. Looking into the segmentation file provided by John, with the idea of possibly "correlating" this information with the clustering exercises already performed, and maybe even somehow assisting some new clustering variation.

Week 8

1. Brent Jones
 - a. team mtg.

Week 7

1. Carlos Rueda
 - a. Group meeting. Some paper reading.

Week 6

1. Danelle Cline
 - a. Applied and was approved by AWS open data program to freely share acoustic data - all 95 TB!
 - b. Assisting student Daniel DeLeon with embedded TensorRT code project. Having success using Google Colab+GCP for this effort.
 - c. Was invited and joined the [OrcaSlack channel](#)
 - d. Testing Keras inference speed improvements
2. Kevin Gomes
 - a. Starting to scope the migration of the backend Alexa skill code from NodeJS 8 to 10 as AWS is deprecating NodeJS 8 very soon.
3. Brent Jones
 - a. Working on getting Sam Urmy's click detection software running in Julia. Sam is leaving so I will be taking over daily processing of Hydrophone.

Week 5

1. Danelle Cline
 - a. Teleconference with AWS about open data sharing program
 - b. Shared notebook examples and helped setup notebook in GCP for Cal Poly student Daniel DeLeon to train models.

- c. Cleaned up wandb model reports organizing by project number instead of name