HALLO Data Portal

Summary

The HALLO (Humans and ALgorithms Listening for Orcas) project has compiled several acoustic datasets containing orca (*Orcinus orca*) vocalizations as part of its efforts to build better automated acoustic detectors and classifiers to support research and conservation initiatives targeting orcas.

In addition to compiling relevant acoustic recordings, the HALLO team also added value to these datasets creating or improving annotations to several levels of detail, from species-level presence to ecotype, pod and call type classification when possible.

The HALLO data portal will make these valuable data resources freely available to the wider research community in order to foster continued developments in automated detection and classification of orca vocalizations.

Project Charter

Team

Core members:

(Work on the project though its entire duration)

Fabio Frazao (Project Manager)

Jack Zhang (Web Developer)

Auxiliary members:

(Contribute with occasional tasks, helping the core members with specific tasks that fall under the auxiliary member's expertise/role)

Sadman Sakib (Data Analyst)

Jason Newport (Data Manager)

Jillian Anderson (Server Admin)

1. Goals

- Develop a web portal to enable registered users to download pre-bundled HALLO datasets (audio files + annotations + metadata) and data managers to add new datasets in the future.
- Deploy the web portal on the compute canada-hosted virtual machine (orca-vm) and have it live by March 31st 2022
- Write documentation providing instructions on how each role (visitor, registered user, data manager, admin) can use the data portal, as well as instructions on how to deploy it.

2. Scope

The primary intention of these datasets is to support machine learning efforts.

The data portal **does not** provide a reference call catalogue. This is the subject of a separate project within HALLO.

The following datasets will be available as bundles:

1-Coarse annotations

[Presence level only, not verified for false negatives, possibly non-annotated calls]

- JASCO Robert's Bank dataset
- JASCO Boundary Pass dataset
- ONC Barkley Canyon node dataset (subset containing orcas

2- Detailed annotations

[Presence, ecotype, pod and call-type levels whenever possible; fully inspected by a bioacoustician; confirmed negatives]

- JASCO Roberts' Bank
- JASCO Boundary Pass
- ONC several nodes

Users will be presented with with summary pages describing the datasets available:

This information will include:

- Time period covered
- Location name
- Coordinates
- Depth
- Instrument used
- Audio format
- Sampling rate
- Dataset size (in mb and hours of recording)
- Annotation level (presence only or if it includes ecotype, pod and call type information
- Number of annotations, detailing how many annotations per class

Filtering:

Users will be able to filter datasets by any of the criteria in the list above

Visualizations

Some of the dataset metadata will be presented in interactive visualizations. Namely:

- A map displaying the location of datasets

- Histograms showing the number of annotations per class at different levels (Species presence, ecotype, pod, call type)

3.Deliverables

- A repository containing the codebase for the data portal
- A live instance of the portal deployed on the compute canada cloud virtual machine (orca-vm)
- Documentation detailing how to deploy and use the data portal

4. Success criteria

1. A live instance will be running on the orca-vm instance by March 31st 2022
2. A test session will be conducted on the live instance and users must be able to:
    - Successfully download datasets
    - Successfully add a new dataset
3. The code repository will be shared with the stakeholders and approved (by them or someone they appoint)[Suggestion: Dr. Steven Bergner, SFU)
4. The documentation will be shared with stakeholders and approved (by them or someone they appoint) [Suggestion: Dr. Steven Bergner, SFU)

    Suggested test group:

    - Dr. Scott Veirs

    - Dr. Val Veirs

    - Dr. Ruth Joy

5. Project references

(Provide guidelines for the project deliverables and work with the project manager to define requirements)

Dr. Ruth Joy (SFU)

Dr. Oliver Kirsebom (Dal)