

# ORCHIVE: Digitizing and Analyzing Orca Vocalizations

**George Tzanetakis & Mathieu Lagrange**

Department of Computer Science  
University of Victoria, Canada  
{gtzan, lagrange}@uvic.ca

**Paul Spong & Helena Symonds**

Orcalab, Hanson Island, BC, Canada

## Abstract

This paper describes the process of creating a large digital archive of killer whale or orca vocalizations. The goal of the project is to digitize approximately 20000 hours of existing analog recordings of these vocalizations in order to facilitate access to researchers internationally. We are also developing tools to assist content-based access and retrieval over this large digital audio archive. After describing the logistics of the digitization process we describe algorithms for denoising the vocalizations and for segmenting the recordings into regions of interest. It is our hope that the creation of this archive and the associated tools will lead to better understanding of the acoustic communications of Orca communities worldwide.

## Introduction

The fish eating killer whales or orcas, *Orcinus Orca*, that are the focus of this project live in the coastal waters of the northeastern Pacific Ocean. They are termed *residents* and live in the most stable groups documented among mammals. Resident orcas emit a variety of vocalizations including echolocation clicks, tonal whistles, and pulsed calls (Ford et al., 2000). The Northern Resident Community consists of more than 200 individually known orcas in three acoustic clans. It is regularly found in the study area of the Johnstone Strait and the adjacent waters off Vancouver Island, British Columbia from July to October.

The goal of the Orchive project is to digitize acoustic data collected over a period of 36 years using a variety of analog media at the research station OrcaLab on Hanson Island which is located centrally in the study area. Currently we have approximately 20000 hours of analog recordings mostly in high quality audio cassettes. In addition to the digitization effort which is underway we are developing algorithms and software tools to facilitate access and retrieval to this large audio collection. The size of this collection makes access and retrieval especially challenging (for example it would take approximately 2.2 years of continuous listening to cover the entire archive). Therefore the developed algorithms and tools are essential for effective long term studies employing acoustic techniques. This project is just beginning but we believe it provides many opportunities and challenges related to large scale semantic access in a non-typical application scenario. We are looking forward to receiving feedback from other researchers in information access and retrieval regarding this project. Finally we hope that in the future researchers from around the world will be able to access this repository and use the developed tools to improve understanding of acoustic communication of orcas.

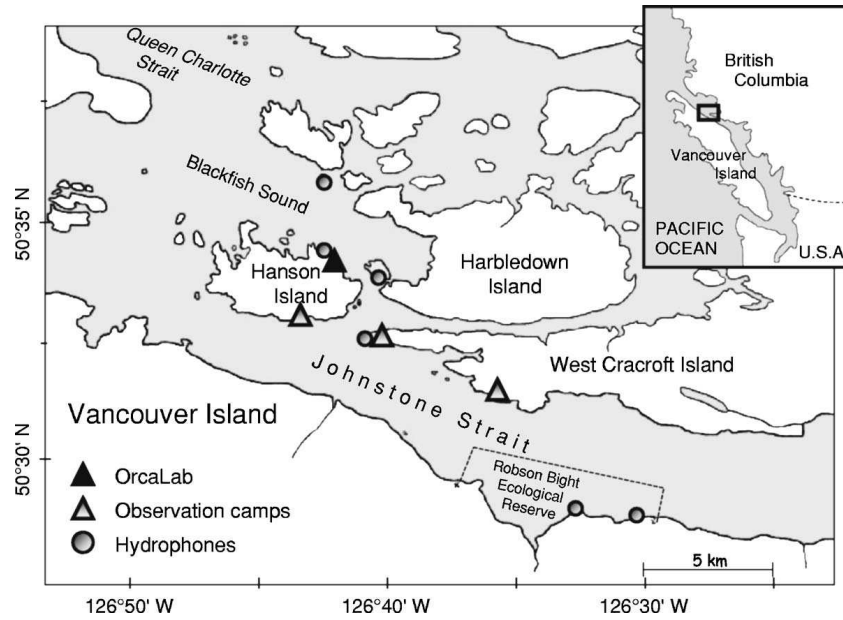


Figure 1: Summer core area of the Northern Resident community with land-based observation sites and the OrcaLab hydrophone network.

### The recordings

The acoustic data has been collected with a network of up to six radio-transmitting custom-made hydrophone stations (overall system frequency response 10 Hz to 15 kHz) monitoring the underwater acoustic environment of the area continuously, 24 hours a day and year around. Whenever whales are vocal the mixed output of radio receivers tuned to the specific frequencies of the remote transmitters is recorded on a two-channel audio cassette recorder. Use of mixer controls allows distinction of hydrophone stations and thus basic tracking of group movements. In addition to the acoustic data corresponding information is recorded in logs that are associated with specific tapes. This information is based on acoustic and visual data collected by volunteers, other independent researchers and whale watch operators. It includes various types of information such as number and identity of individuals, group composition, group cohesion, direction of movement and behavioral state (travel, motionless, forage and socialize). The logs also record the mixer settings.

The majority of the audio recordings consist of three broad classes of audio signals: background noise caused mainly by the hydrophones, boats etc, background noise with orca vocalizations and voice over sections where the observer that started the recording is talking about the details of the particular recording.

### Digitization

The logistics of digitizing all these analog audio tapes are challenging. We plan to record the 20000 hours of audio to uncompressed digital audio in stereo with a sampling rate of 44100 Hz, and 16-bit dynamic range. This results in approximately 1 GB / hour of audio so we will eventually require storage for 20 Terrabytes (TB) of data.

After taking into consideration various tradeoffs between time, cost, robustness, power consumption (important for Orcalab which is not connected to a power grid) and other factors we decided to start the pilot phase of the project using two digitization stations. One of the stations is located at the University of Victoria and the other is located at Orcalab. Once the pilot phase is completed and any potential problems are worked out we will be able to scale the digitization effort by purchasing more digitizing stations. Each digitizing station consists of the following components: 1 Apple Mac Mini small-factor desktop computer, 2 dual tape Tascam 322 cassette players and 1 Tascam FW1804 multichannel firewire audio interface.

Each digitization station is capable of recording 4 stereo analog audio cassettes (8 channels) simultaneously at 44100 Hz. Although it is possible with more specialized hardware to record more channels simultaneously into 1 computer we have that 8 channels is best in terms of cost effectiveness and robustness. Custom software has been written to simplify the process of digitization requiring minimum human involvement (just manual loading of the cassettes and pressing the play buttons). Assuming 6 hours of digitization per day one station is capable of processing 16 tapes per day it would take approximately 3.5 years to digitize the entire archive. More digitization speed up the process linearly. Currently we have digitized approximately 200 tapes corresponding to 1% of the total archive. Digitizing this data has enabled us to iron out problem with throughput and the software as well as provide a test-bed for the tools described in the following sections. For storage we are currently using a combination of three storage devices: high-quality DVD+R are used for long term archiving as they have a much longer life than regular DVDs, 360 Gigabyte external hard drives are used for communication of data between the two locations and temporary storage, and a 7 Terabyte Apple XServe-Raid that we plan to scale up eventually is used for storing the overall archive and for data processing.

## **Denoising**

In most of the recordings the background noise level is very high. This is caused by water movement, rain drops, passing boats and underwater acoustic transmission. Therefore denoising is more challenging than standard audio recordings using regular microphones. Standard denoising algorithms require the user to provide a recording of the background noise, calculate its statistics and subsequently use these statistics to filter out the corresponding frequencies, see Vaseghi et al (1992) for further references. In many of the orca vocalizations standard denoising approaches fail as in the case of boat noise with moving frequencies dues to changes of the engine speed. On the other hand orca vocalizations are optimized for transmission in the noise underwater environment and therefore exhibit strong harmonic peaks. In order to separate the orca vocalizations from the background noise we utilize a new data-driven algorithm for simultaneous partial tracking and sound source formation proposed in Lagrange and Tzanetakis (2006). The algorithm is inspired by ideas in Computational Auditory Scene Analysis (Bregman, 1990).

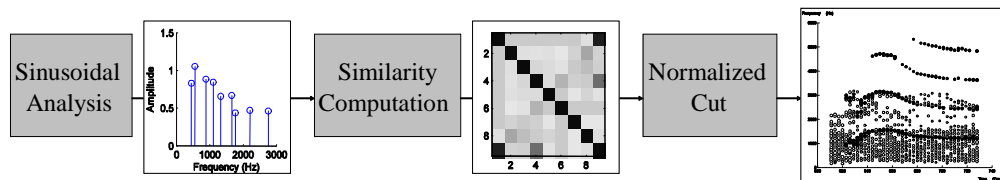


Figure 2: Block-Diagram of the audio processing unit used for denoising orca vocalizations.

Frequency bands with high energy are identified and similarities among them are computed. A graph algorithm called the *normalized cut* is then used in order to organize those prominent components according to several constraints. The modeling of human vocalization using time-varying sinusoids is a well-known and validated approach. Using such a model for orca vocalizations is also possible but lead to identification problems due to the non stationary nature of these calls. Preliminary experiments show that this algorithm is able to deal with this issue by optimizing continuity and harmonicity constraints globally in a larger time scale.

### Classification and Segmentation

Locating a particular segment of interest in a long monolithic audio recording can be various tedious as users have to listen to a lot of irrelevant parts until they can locate what they are looking for. Even though visualizations such as spectrograms can provide some assistance they still require a lot of manual effort. In this section we describe some initial proof-of-concept experiments for the automatic classification and segmentation of the orca recordings.

We have built a classifier for the automatic detection of the three main typos of audio encountered in the recordings namely: voice over, background noise and background noise with orca vocalizations, see middle of Figure 3. Initial evaluations using either artificial neural networks or support vector machines are very encouraging. By using a subset of the features proposed in Tzanetakis and Cook (2002) for musical genre classification we are able to correctly classify audio frames of 1 second with 95% accuracy. These initial experiments were done using 10 minutes of audio for training and 10 minutes for testing. The training data and testing data came from different analog cassettes. We are currently working on using larger datasets for training and testing.

The most common vocalizations are “discrete calls” which are highly stereotyped pulsed calls that can be divided into distinct call types (Ford, 1989). Studies have showed that “pods” which are social units comprised for one or more closely related matriline have unique vocal repertoires of 7 to 17 discrete call types. Even we still haven’t developed algorithms to classify calls into types we utilize a variation of the segmentation methodology proposed in Tzanetakis and Cook (2000) to identify the onsets of these “discrete calls”. That way, researchers can directly skip forward and backward through the recording based on semantic units (the “discrete calls”) rather than arbitrary units selected by the zoom level of the audio editing software.

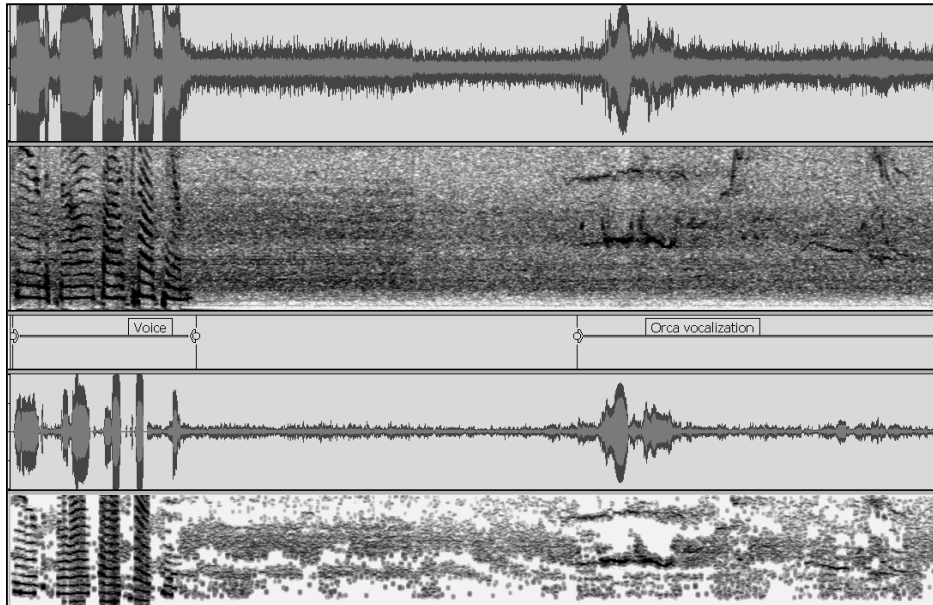


Figure 3: Automatic labeling and denoising of Orca recordings using Marsyas tools. The bottom two time domain and spectral plots correspond to the denoised signals.

### Discussion and Future Work

All the software and tools used in the Orchi project are developed using Marsyas (<http://marsyas.sourceforge.net>) which is a free software framework for audio analysis, retrieval and synthesis. It is important to emphasize that our goal is to provide tools and support to assist researchers in understanding orca vocalizations rather than replacing the human element in the process. The digitization effort is under way and we believe we are ready to scale to more stations and larger archives. The Orchi project is just starting so there is a lot of room for future work. Future directions we plan to explore include: automatic identification of individual calls using supervised learning methods, classification to pod and matriline, similarity detection to identify calls across time. Another goal is the development of a continuous monitoring system that automatically detects when orcas vocalize and starts direct digital recording. Finally we plan to provide many of our access tools as web services so that researchers can access directly the parts of the signal they are interested in without having to download the entire recording. As an example scenario a researcher might request all instances of N2 (a particular type of call) over the period of 1995-1997 with the background noise removed.

### References

- Bregman A. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound. Cambridge, Massachusetts: MIT Press
- Ford, J.K.B. (1989) Acoustic Behavior of Resident Killer Whales (*Orcinus Orca*) off Vancouver Island, British Columbia. *Canadian Journal of Zoology* 64, 727-745.
- Ford, J.K. B., Ellis, G. M., and Balcomb, K.C. (2000). Killer Whales: The natural history and genealogy of *Orcinus Orca* in British Columbia and Washington, 2nd ed (UBC, Vancouver)
- Lagrange, M., and Tzanetakis, G. (2006) Sound Source Formation and Tracking using the Normalized Cut. Submitted to "Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)
- Vaseghi S. V. et al (1992) Restoration of old gramophone recordings "Journal of the Audio Engineering Society", vol. 40, n° 10, Oct. 1992.
- Tzanetakis, G. and P. Cook (2002) Musical Genre Classification of Audio Signals, *IEEE Transactions on Speech and Audio Processin*, 10(5)
- Tzanetakis, G. and P. Cook (2000) Multi-Feature Audio Segmentation for Browsing and Annotation" In Proc. IEEE Workshop on Applications of Signals Processing to Audio and Acoustics (WASPAA).