CATEGORIZATION OF MARINE ORGANISMS DETECTED BY SEAL-BORNE ACTIVE ACOUSTIC LOGGERS (MICROSONARS) USING MACHINE LEARNING METHODS.

Maria Isabel dos Santos Barros
M2 Marine Physics - IUEM

Advisors:
Mickaël Coustaty, L3i/La Rochelle Université
Tiphaine Jeanniard du Dot, (CEBC, CNRS)

September 3, 2021

Acknowledgements

I would like to thank my supervisor, Mickaël Coustaty, for his guidance during this internship. Also, to Jeanniard-du-Dot, for co-supervising this work, for the kind words of incentive, for the disposition to help me understand the subject and for reading and correcting this report.

To Pauline Goulet, that since the beginning helped me understand the processing of sonar data, for kindly answering all my doubts and for evaluating the image processing method developed in this work. Also, to Mathilde Chevallay, who also evaluated our image processing method and provided import expert insights of it. To Baptiste Picard and Martin Tournier, the CEBC staff that helped my with the sonar tag and GPS database.

To Pierre Tandeo, for letting me know about this project, for the support in the image classification part of this project and for the guidance during all my Masters second year.

To the awesome friends I made here in La Rochelle, Guilherme, Layssa, Larissa, João and Débora, for making my stay a lot happier, for all the adventures and the great beach days.

To Wilton Aguiar, my dear friend and fellow oceanographer, for the help with the maps and the writing of the report, and your support and for your always positive disposition.

Lastly, to my mom and dad, Sueli and Lucas, and to my sister Beatriz, who supported me during all this Masters, even at distance cheered me on difficult times and helped me keeping focused on my goals in these last two years.

Contents

Chapter 1

Introduction

The world's oceans stored more than 90% of the energy accumulated in the climate system between 1971 and 2010, and surface waters consequently warmed by $0.11^{\circ}C$ per decade during this period(*Portner et al.*, 2019). Among them, the Southern Ocean, surrounding the Antarctic continent and connecting the Indian, Pacific and Atlantic Oceans through the Antarctic Circumpolar Current (*Gille*, 2002), plays a major role in global climate dynamics; and the increase in its water temperatures in the upper 2000m depths represents up to 52% of the increase of ocean heat content from 2005-2017, leading to significant consequences for large-scale biogeochemical cycles (*Portner et al.*, 2019). These changes impact the functioning of the ecosystem and the distribution of marine organisms from zooplankton to fish populations and in fine top predators. These top predators such as birds and marine mammals integrate the effects of changes throughout their food web, and are therefore considered bioindicators of these changes. Biogeochemical changes affect their behavior, fishing efficiency, body condition, health, reproductive success and survival, through the decrease in prey abundance and accessibility (*Younger et al.*, 2016).

The mesopelagic part of the marine ecosystem dynamics has been understudied, due to lack of technological tools to do so in conjunction with oceanographic parameters. This community is important as preys to marine mammals, marine birds, and large pelagic fish (*St. John et al.*, 2016). To study prey is challenging because it is distributed in patches that change fast in the marine environment, making it difficult to track (*Benoit-Bird and Au*, 2003). One of the main ways to investigate prey distribution in the ocean is to study the behavior of their predators in conjunction with habitat characteristics(e.g. *Wisniewska et al.* (2016), *Goulet* (2020)).

In the Southern Ocean, one of the most abundant and widespread top predator, and one of the main consumers of marine resources in the area, is Southern elephant seal (*Mirounga leonina*) the largest member of the Phocidae family (*Guinet et al.*, 1996). These seals spend 10 months per year at sea to forage, and go back to land on sub-Antarctic territories twice a year, to breed (October) and to moult (February) (*Field et al.*, 2004). When at sea, elephant seals travel several thousand kilometers, during which they dive continuously to the 300-500m depth zone on average, but to a maximum depth of 2000m and duration of more than an hour. Elephant seals mostly feed on myctophids (*McIntyre et al.*, 2010), small energy-dense and very abundant mesopelagic fish, or on high trophic level bathypelagic prey such as cephalopods, and tend to follow migration of their prey by diving deeper during the day than at night (*Hindell et al.*, 1991). These top predators forage in different environments, depending on their age and sex (*Campagna et al.*, 2006).

However, the remote location of these animals foraging environment make them difficult to directly study, and particularly in conjunction to their environment. Biologging, the use of tags and other devices to record data in animals, is an important tool to study the marine environment (*Ropert-Coudert et al.*, 2010). These sensors attached to animals is a solution to the challenges of studying the distribution of organisms in ecological studies with in *in situ* data. They have large applicability to obtain this information from areas that cannot easily be accessed using vessels or even autonomous floats. The first works that used tagging in bigger marine mammals mainly investigated diving behaviours (e.g., *McConnell et al.* (1992) and *Fedak et al.* (2001)). Most recently, the development of technology of microprocessors enabled the application of sensors of temperature, depth, bio-luminescence, salinity, cameras and GPS position in marine animals *Block* (2005). These types of equipment have large applicability to study animals populations and its influence on individuals, but still lacks the capacity to evaluate mid-trophic level organism assemblages in direct relation to the predators and their environment.

To bridge this knowledge gap, an innovative newly developed biologging tag called microsonar was recently developed *Goulet et al.* (2019)). This microsonar mimicks echolocation capabilities of cetaceans and provide information on prey fields right in front of the animal's head while it is diving to find prey. The microsonar logs high-resolution (25Hz) echosounding data for months at a time when animals are at sea and generate large amount of data that would be extremely and prohibitively time-consuming to visually process in their entirety. For each animal borne tag, more than 100.000 sonar images can be generated in a single deployment. To visually inspect each echogram in order to assess the presence of prey, their type, and other relevant ecological information, would take a vast amount of time. Also, the interpretation of echograms is a subjective process, which needs validation and inter-comparison among experts (*Wisniewska et al.*, 2016).

It is consequently necessary to develop automated ways to interpret these echograms and the use machine learning capabilities to process echosounding images and train models to automatically classify our images of interest. Typical methods of image classification may involve labelling connected components (*Dillencourt et al.*, 1992). In this method a threshold is used to segment a gray image, that is then binarized, i. e. this image is converted to black and white. Afterwards segments of connected pixels, the objects, are extracted and then characterized. Finally, classification models can label these objects based on their set of features, which are relevant information about the image, such as color, corners, edge, texture, etc (*Hassaballah and Awad*, 2016). The classification methods can be supervised, where objects are already labeled. Conversely, they can be unsupervised, where the labels are not yet known and the model will create the classes of objects based on their properties, for example their color information (*Dhaware and Wanjale*, 2016).

In supervised machine learning there is a pre-processing stage, where noisy data, empty data and outliers are removed. Then the features of this data is selected, given attention to the representativity of those features and avoiding redundancy. Part of this processed data is manually classified, where labels will be assigned to this data, the training dataset. Algorithms for data classification are is chosen and they can be evaluated by splitting the training dataset in, for example, two thirds to 70% to train the algorithms and the remaining data is used to estimate their performance. Fine tuning of parameters is performed. Once satisfactory results are achieved the model algorithm is used to classify

the datasets of interest (*Kotsiantis et al.*, 2007).

Common supervised algorithms used in image classification are K nearest neighbors (KNN) and Support Vector Machines (SVM). The first method uses the label information of the k adjacent data points to classify an image (*Kim*[1] *et al.*, 2012). Its advantages are the simplicity, effectiveness, as well as low computation cost, once the training step comprises of assigning the class to data based on the classes of their neighbors. The disadvantages are that the method's high sensitivity to the number of neighbors and to noisy regions, that may cause mislabeling, and to the size of each class, as bigger classes will influence more the adjacent data. The second method, SVM, transform the training data into a high dimension space in order to linearly separate the data into classes, maximizing the margin between the classification plane and the data that are in the boundaries of the classes (*Chandra and Bedi*, 2018), called support vectors. It performs well with a large number of features, and continuous features (*Kotsiantis et al.*, 2007), whereas this method does not work well when the classes cannot be linearly separated, and the data needs to be mapped in a higher dimensional space. Another disadvantage is that this method higher complexity, compared to KNN, leads to higher computational cost when training the model.

The performance of supervised methods will highly depend on the quality of the training dataset, as it is build manually and can be associated with errors (*Pasolli et al.*, 2013). Sonar images are difficult to process manually, even by experts, due to presence of back-scattering of the sound signal from different organisms and particles, resulting in noisy images. Consequently, in order to be able to automatically interpret the echograms and retrieve relevant information as the presence/absence of prey, prey type and other relevant information, it is necessary to learn the most suitable pre-processing methods, image features and classification.

1.1 Objective

The main objective of this project is to develop a methodology for processing micro-sonar images based on image processing and machine learning methods to automatically and efficiently extract relevant information from the obatined echograms. More specifically, we aim to:

1. isolate the portion of the dataset where relevant predator-targeted prey encounter signals are present in the echograms,

2. classify these extracted signals into prey type (single vs schooling prey, passive versus active), and categories of acoustic signal size and intensity,

3. identify the vibrissae deployment events used by elephant seals to locate their prey visible on the echograms.

1.2 Outline of the report

Section 2 presents some theoretic background on echosounder as well as the description of the sonar tag and the data used in this work. In section 3 we present the processing of sonar

images developed in this work. The algorithm used to identify the main image features, as well as the presence/absence of prey is presented in details, as its implementation constituted a significant part of my work. In section 4, we present the applicability of the pre-processing method, as well as the performance of the classification algorithms we chose. We also present a census of the deployment of whiskers by southern elephant seals. Finally, a discussion with some concluding remarks and perspectives are presented section 5.

Chapter 2

Context

This work used the data analysis in sonar tags attached to elephant seals. This tag was developed in the Sea Mammal Research Unit, St-Andrews. All data recorded is stored in a Centre d'Etudes Biologiques de Chizé (CEBC) server. From the records, the part that would be analyzed was separated into smaller datasets, accounting the time when the seal was apparently pursuing a prey, as the focus of the study is the predator-prey encounters. Part of this dataset was already analyzed, for other works. Three people from the CEBC, one of the initial microsonar developers and two students, worked on the training dataset that was used to train and test our classification method in this work. From that we used the infrastructure of the Laboratoire Informatique, Image et Interaction (L3i) at La Rochelle University to perform all the analysis. This work was done remotely, but had consultation with members of the CEBC and direct contact with the people that worked on this dataset.

2.1 Functioning of sonars

Sonar is an equipment used to detect and observe objects underwater (*Simmonds and MacLennan*, 2008). An echosounder is an active type of sonar that emits acoustic beams and detects the returning signal, i.e. echo, from various objects, called targets. There are various applications to echosounders: fisheries, seabed mapping, defence (submarine detection). The echosounders have two principal components, a transmitter that produces electrical energy at a given frequency, depending on the object to be detected, and a transducer, that converts this electrical signal into an acoustic wave in frontal beams(Fig. 2.1). The beam width in echosounders goes from $5-15$ deg. This pulse travels in the water, hit a target, that will scatter part of this beam and the other part will be detected back by the transducer. The transducer transforms back the sound wave into electric pulse that is the received signal. The time that the sound takes to get to the targets and get back to the echosounder is used to calculate the distance of the target to the sonar. After some time specified by the operator, the transducer sends another pulse, repeating the process. The pulse emitted by the transducer is the source level (SL). Then there is transmission loss (TL) due to the medium, that is calculated for the incident and the returning signal. The intensity that is detected by the echosounder is the target strength (TS), the ratio of the incident sound wave that hit the target and is backscattered (*Simmonds and MacLennan*, 2008). The target strength will strongly depend on the temperature and is represented in logarithmic scale. Then the final signal measured by the sensor will be the echo level (EL)

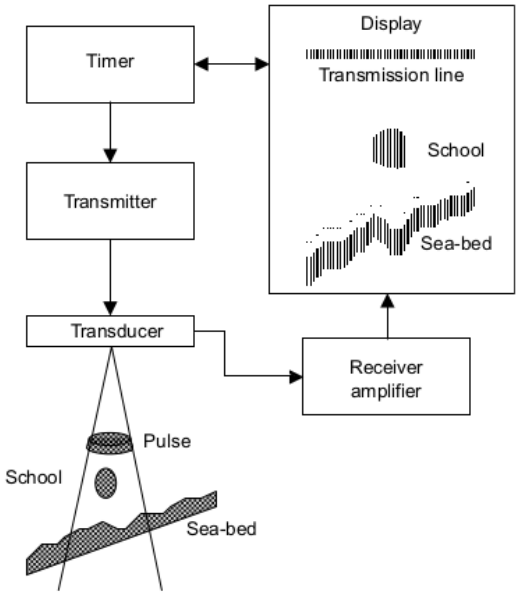

Figure 2.1: Schematics of an echosounder. The electric signal is transformed in sound waves by the transducer, hit the fish school and the seabed and receives the echo. On the right, the echos are displayed in echograms.

calculated as:

$$TS = 10log_{10}(\frac{\sigma_{sp}}{4\pi}) \tag{2.1}$$

$$EL = SL - 2TL + TS \tag{2.2}$$

whre σ_{sp} is the strength of back scattering. Finally, the signal from targets is presented in signal-to-noise ratio (SNR), that is the difference between the power of the signal (EL) and the background noise, also given in decibels. The noise can be generated by smaller particles, such as the phytoplankton, and ocean turbulence.

2.2 Field work procedure

Two post-breeding female southern elephant seals (*Mirouga leonina*) were captured in Patagonia, Argentina (Peninsula Valdés, 42°57'S - 63°59'W, hereafter named PV1 and PV2) in October of 2018, and three were captured on the Kerguelen Islands (49°20'S - 70°20'E, hereafter named KER1, KER2 and KER3) in October of 2018 (Fig. 2.2). They were captured on land with a canvas hood and anesthetized with an intravenous injection of Zooletil 100©($\sim 0.8mL$ per 100 kg of estimated body weight). Once anesthetized, they were equipped with a micro-sonar tag on the head (DTAG, Mark Johnson, SMRU, 95 x 55 x 37 mm, 200 g in air), in addition to a SPOT-293 Argos tag attached to the neck (Wildlife Computers, USA, 72 x 54 x 24 mm, 119 g in air), and a CTD-SRDL tag (Conductivity Temperature Depth - Satellite Relayed Data Logger) on the back (Sea Mammal Research Unit, St-Andrews, UK, 115 x 100 x 40 mm, 680 g in air).

The tags were glued to the fur using a fast-setting epoxy glue with a low exothermic reaction. They were recaptured in December/January when the animals returned to land to molt, and the tags recovered using the same anesthesia procedure. This

project and all the procedures used on animals were approved by the Ethics Committee Anses/ENVA/UPEC (permit number 19-040 21375) and the TAAF Polar Environment Committee (no permit number provided). The 3 tags have a cumulative weight of less than 0.5% of the mass of the animals and as such have no detectable consequences on their foraging performances or probability of survival (*McMahon et al.*, 2008).

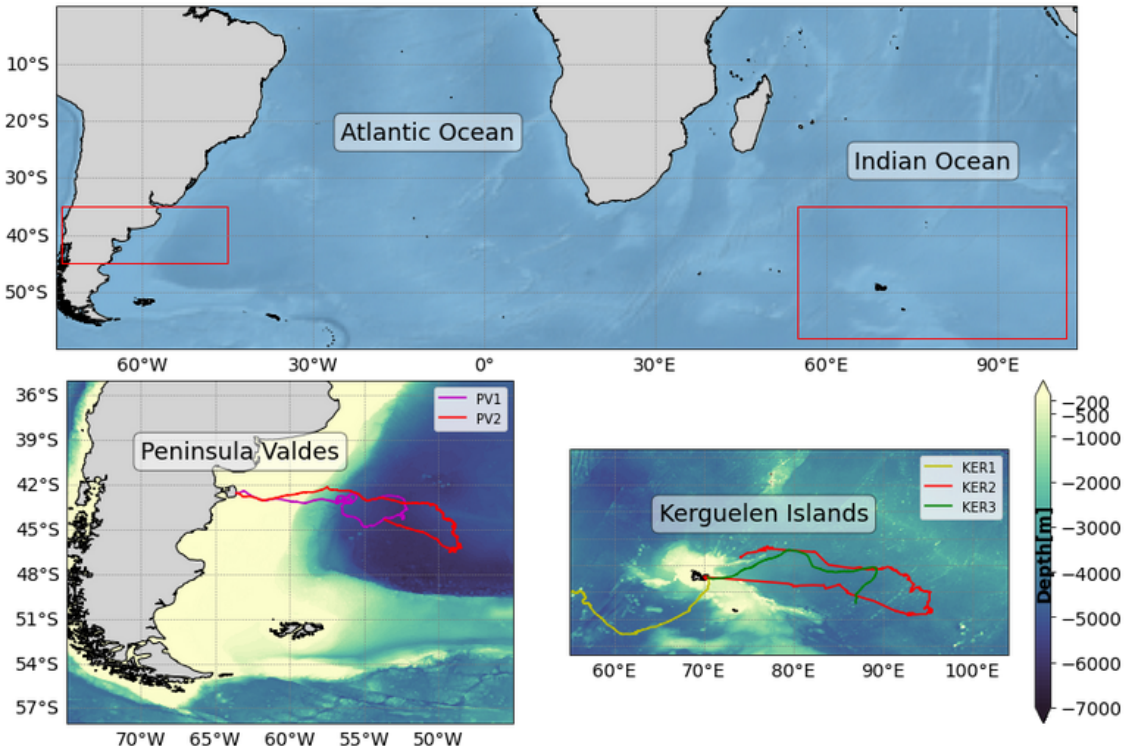

Figure 2.2: Map with the regions where the seals were tagged. The two lower panels are a zoom of the tagging site in Argentina on the left and the region of the Kerguelen Islands on the right, both present bathymetry contours. The The colored lines are the tracks of the seals.

2.3 Microsonar data

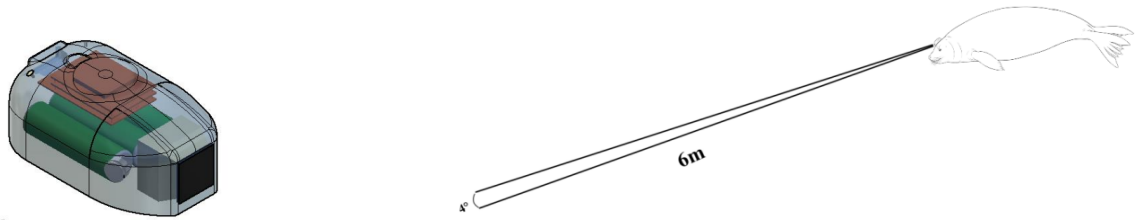

Figure 2.3: On the left: image of the sonar tag. On the right: schematics of the beam emitted by the microsonar.

The microsonar (DTAG) recorded tri-axial acceleration at 200 Hz, tri-axial magnetic field at 50 Hz, dive depth at 50 Hz, geographic position (GPS location every 5 min when the animal's head was submerged) (*Goulet et al.*, 2019) in addition to echosounding

of the environment in front of the animals. This active micro-sonar part of the DTAG emitted 10 s sounds with a center frequency of 1.5 MHz (source level of about 187 dB re 1 Pa at 1 m), at 12.5 (in 2018) or 25 (in 2019) pings per second. The sound was emitted and received via a transducer with an effective beam width of about 3.4ř (based on the 3 dB decay). The echoes generated by the insonified particles/organisms were sampled at 192 kHz for intervals of 8 ms after each ping produced. The center frequency of sound emission was chosen to be above the hearing limit of elephant seals (42 kHz, *Kastak and Schusterman* (1999)) to avoid any sensor-induced disturbances to the animals. *Goulet et al.* (2019) noted that potential sidebands due to the shape of the signal could still interfere with hearing capacities of the seals but noticed no behavioural differences during the periods of time when the sonar was active compared to silent.
The micro-sonar part of the DTAG was programmed to switch on only at depths below 20m and to work on a 24-hour on/off cycle for the 2018 deployments and 24 hours every 72 hours in 2019, to extend battery life and thus the recording time as much as possible (between 16 and 22 cumulative days for each seal). All other parameters collected by the tag were sampled continuously during the post-reproduction foraging trips that lasted 48 - 63 days until the memory of the tag became saturated, and stopped recording. All data remained stored in the tag and was recovered upon return of the equipped seals on land.

Chapter 3

Materials & Methods

3.1 Processing and selection of echograms

The raw acoustic data from the micro-sonar are represented as echograms composed of successive echos from the insonified particles back from the initial emitted pings. An echogram of one second is thus composed of 12.5 or 25 pings (in 2018 or 2019). Each pixel of these pings have a vertical resolution of 3.9 mm, and correspond to the value of the intensity of the 12.5 or 25 pings (in 2018 or 2019). However, the SNR providing more of a qualitative than quantitative information on the transmission, the background noise was calculated as the 5^{th} percentile of the signal and was subtracted to it to obtain the Echo to Noise Ratio (ENR) in analyzed echograms. Despite the single-beam and high frequency nature of the micro-sonar, the acoustic size of insonified organisms was found proportional to their actual size in laboratory experiments in controlled basins, albeit with an effect of the distance, orientation and position of the target in the sonar beam (*Petiteau*, 2020). However, the intensity of the echo received by the transducer for each insonified organism ultimately depends on its backscatter power or target strength, which in turn depends on its density, size, orientation to the acoustic beam, and the presence or absence of a swim bladder (i.e. air in it). The range of colours in the echograms consequently indicates the different intensity of return signal, where yellows to reds represent stronger echos, and blues weaker ones.

As microsonars recorded continuously but we are only interested in the images showing organisms of interest to the elephant seals, i.e. prey actively chased and/or consumed, the choice of echograms to be analysed were based on information provided by the other sensors of the microsonar, i.e depth and tri-axial acceleration. The tri-axial acceleration can indeed help determine times when the animals attempt to capture prey. Seals increase their chasing behaviours and make an up and forward head movement when capturing prey that are visible on acceleration channels at depth (*Vacquié-Garcia et al.*, 2015). It is considered that when the root-mean- square of the acceleration of the seal surpasses an established threshold ($> 350 m^3/s$), a Prey Capture Attempt (or PrCA) occurs (Fig. 3.1). Consequently, we chose to work only with echograms spanning 5 s before and 2 s after a detected PrCA, i.e. echograms of at least 7 s, most likely to display a prey of interest to the animal.

To illustrate this, Fig. 3.1 A shows the echo of an organism in the range of the diving elephant seal not eliciting any behavioural response from the seal (no jerk signal) nor from the insonified organism (passively getting closer to the female elephant seal with

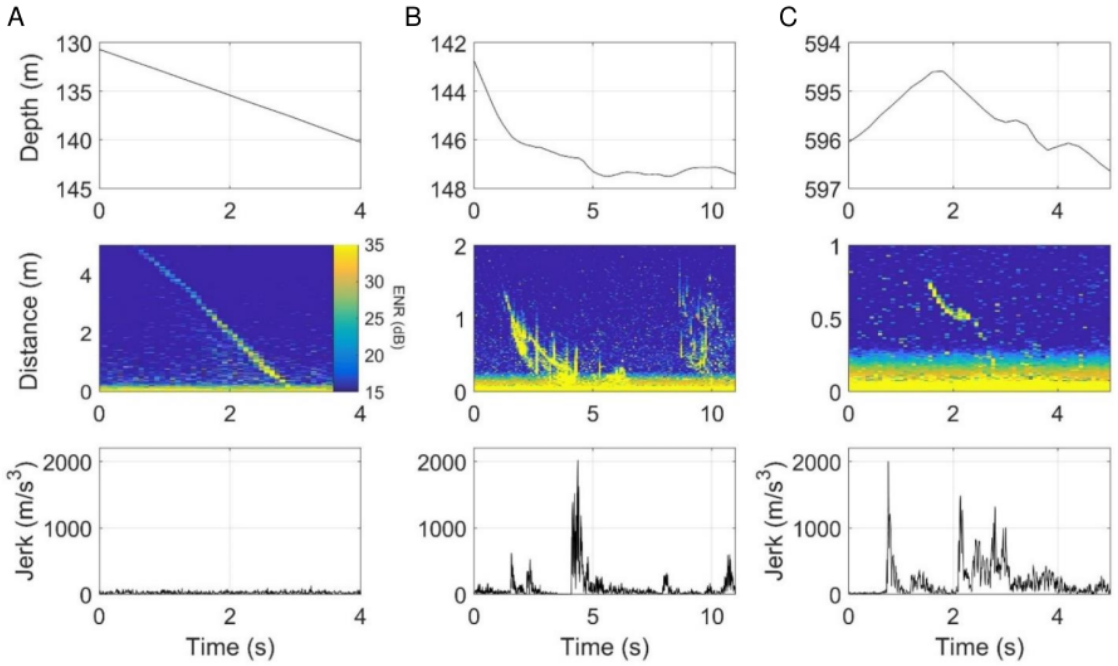

Figure 3.1: Example of simultaneous recordings of dive depths of female southern elephant seals (top row), echograms showing insonified organisms detected by the transducer of the microsonar (middle row), and jerk in acceleration of the head of the animals (lower row) for three different dives and types of encounters (graphs A, B, C). From *Goulet* (2020)

each ping as she is swimming by). Clear jerks in acceleration signals in Fig. 3.1 B and C show that the seals are actively chasing/capturing the prey visible on the echograms. The directional patterns of the insonified targets change with the successive pings indicating an escape response form the single prey (C) or school of prey (B). The 5 seals displayed a combined 100817 echograms from PrCA (20163.4 ± 6413.8 and range: 11152 to 27650 echograms per seal) from which we randomly extracted 38200 echograms to analyze.

3.2 Sonar image processing

The goal of the image processing step is to extract relevant information related to prey, their number, and also the presence or absence of whiskers. The whiskers help the seal detecting preys and can be deployed during a prey capture attempt. As said before, echograms are records of returning signal from particles present in the water column. So in an image we can observe not only the main targets we are interested in, but also the presence of smaller organisms, such as zooplankton, and marine snow, that will not be studied in this work. For this reason, echograms are images that require a complex interpretation. Therefore, all processes, from the decision over what would be considered relevant information to adjustments made to get the final image were done through visual inspection and continuous discussions with an expert. The final echogram that was used for the classification went through a series of steps. In addition experiments were made in each step to reach the optimal processed image.

In general perspective the sonar image processing have followed the steps:

1. Images were separated in main information and background
2. Main features, such as whiskers and bigger components with more intense signal (called traces), were identified
3. These features were reduced to their size, color and sound intensity values
4. These values were used to classify the images into two labels: isolated prey and schooling prey

At first, all image processing was done in echogram images previously visually analysed by experts, that hence created a training dataset containing more than 6000 echograms. This training dataset contains information on echograms from six southern elephant seals, chosen randomly from the main file containing all PrCAs. These experts characterized the images using a custom-built MATLAB toolbox called *sonargame* created by M. Johnson. It generates echograms and allows them to visually interpret the images and infer the presence/absence of prey, their number, whether the prey is reactive to the presence of the seal or not, and their acoustic size, when they are reactive. From this dataset we used 4976 echograms, corresponding to PrCAs from the five seals described previously. After identification of the main features we applied our classification method to the remaining PrCAs (33224).

Once the echograms were selected, the first step in image processing was to perform an image segmentation to separate the important 'prey-related' information from the background. The echogram images have been through pre-processing: as the PrCAs events had varied duration, the images were fixed in 800X1000 size. Secondly, we transformed the echo intensities in decibels to grayscale, for posterior classification. Afterwards we separated the targets in the images from the background - mostly the dark blue area resulting from backscattering of sound waves from smaller particles.

3.2.1 Image segmentation

At first we performed an exploratory analysis of pixels distribution to evaluate the possibility of using an automatic thresholding that could maximize interclass (background and foreground) variance, with the Otsu method (*Otsu*, 1979). For that, the pixels should have a bimodal distribution, that would allow us to separate the image into background and foreground. The Otsu method then aims to find the valley between the two peaks in distribution in grayscale, the threshold. Below we present a histogram of an image in grayscale where the Otsu method would work and a histogram of an echogram, and we can observe the distribution has only one main peak. We consequently discarded this method.

We then attempted was the use a global threshold for all images, using echo intensity values. In their work, *Goulet et al.* (2020) uses this threshold to separate the pixels of probably prey to measure their acoustic size. This method filtered more information than necessary, as we can see in the figure below. In the left we have a raw echogram and the right one is the processed echogram of an schooling prey. We observe the filtering method excluded information and the shape of the target is not recognizable, as one artifact was to convert all components to squares, making the image more difficult to interpret. Another effect is that it might miss smaller targets or preys with weaker intensity and therefore less contrast with the background.

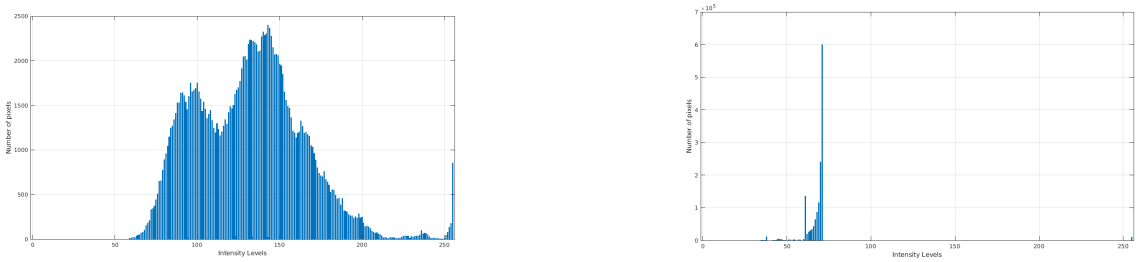

Figure 3.2: Left: Histogram of an sample image in grayscale. Right: histogram of an echogram

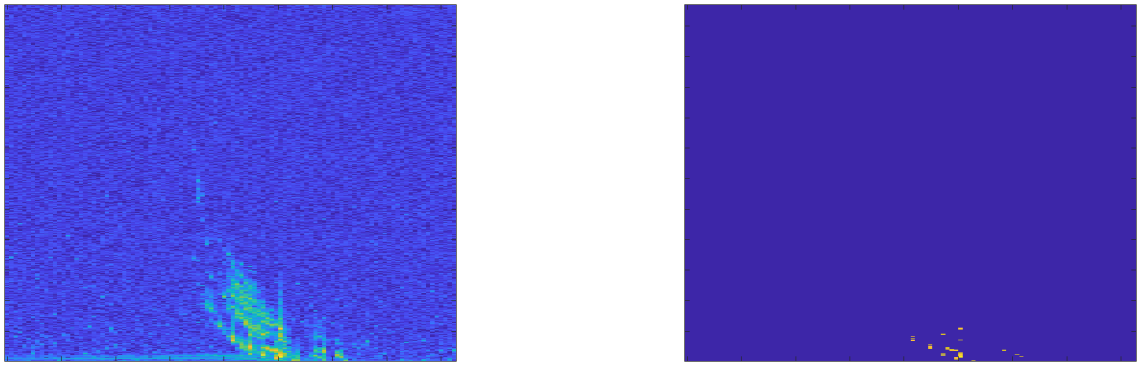

Figure 3.3: Left: Raw image of an echogram. Right: image segmented with the use of a global threshold of 20db less than the strongest signal. All pixels below that value weren't kept.

Finally, we opted for using K- means clustering, adapted to colored images (MATLAB function *imsegkmeans*). K-means is an unsupervised image classification method that separates the data into independent groups in an iterative process (*Na et al.*, 2010). At first we allocate randomly k centers, where k is the number of clusters we must choose previously. Then each data point is assigned to the cluster with the nearest center. After that we calculate the Euclidean distance from each data to the nearest center:

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} \mid x - x_i \mid^2 \tag{3.1}$$

Where E is the squared error between each data point (x) and the center (x_i) for each cluster. Afterwards a we calculate the mean of the data for each cluster, that will then be the new center. These steps are repeated until we reach a local minima. The main objective is to minimize the value of E. In color images (RGB images) each pixel has three coordinates, the Red channel, the Green channel the Blue channel, with intensity values that go from 0 to 255. When applied to color images the k-means method uses the color information of each pixel and the distance between them to assign the pixels to the clusters.

Once more we performed a series of experiments in order to choose the right number of clusters and whether it would be necessary to do the clustering more than once, in order to find the right balance between filtering background noise and not missing important information. We defined two different strategies:

1. Perform the k-means clustering with a number of clusters from 2 to 10 and evaluate which one had the best performance.

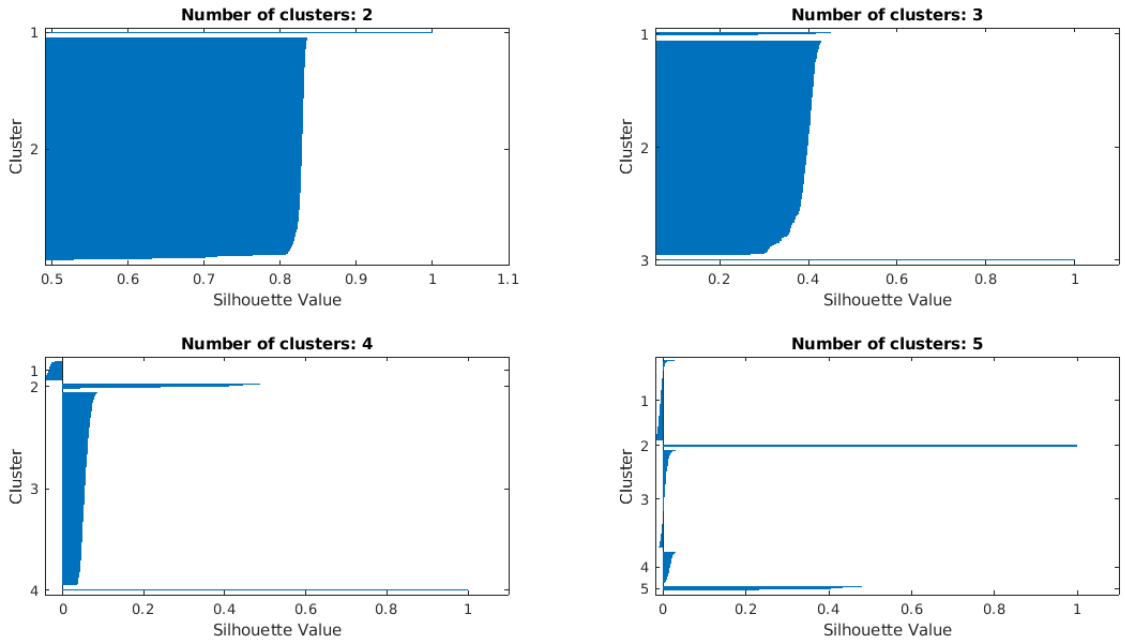

Figure 3.4: Silhouette analysis for the number of clusters from 2 to 6

2. Use $k = 2$ and perform the clustering two times

The first strategy did not provide constant results over the larger dataset. For example, we evaluated the clustering performance with silhouette analysis, to study the separating distance between the clusters. The silhouette has a range of -1 to 1, where numbers close to one indicate that the data in one cluster is far from the other one; and negative numbers mean that the data was possibly assigned to the wrong cluster. The graph in Fig.3.4 shows the silhouette values for different numbers of clusters. The cluster 2 (upper right graph) has the highest silhouette values, being the best choice. Whereas with four and five clusters there are negative silhouette values, and they are not separating correctly the data. Another issue is that many echograms didn't have the same optimum number of clusters, making the interpretation of the results more complex (which clusters had relevant information) and preventing the elaboration of a general scheme to analyze all images. The second strategy had an overall better performance. The use of two clusters may have excluded less information, and thus kept pixels of background noise, but we solved it by performing the clustering another time. This method also had the advantage of being the simplest solution and the one that would exclude relevant information the least. The limitation of this method is that sometimes there was not enough contrast of targets with the background, especially when the seal was close to areas with marine snow or smaller organisms, so the processed images would keep the noisy data Fig. 3.5. As this situation wasn't frequent, we decided to move forward and choose the second strategy to perform image segmentation.

The chosen strategy to perform the clustering of images using k-means algorithm had the following steps:

1. K centroids are chosen randomly, then the pixels are assigned to the cluster with the closest center.

2. The mean intensity value of pixel in the three dimensions is computed for each cluster, becoming the new centroid.

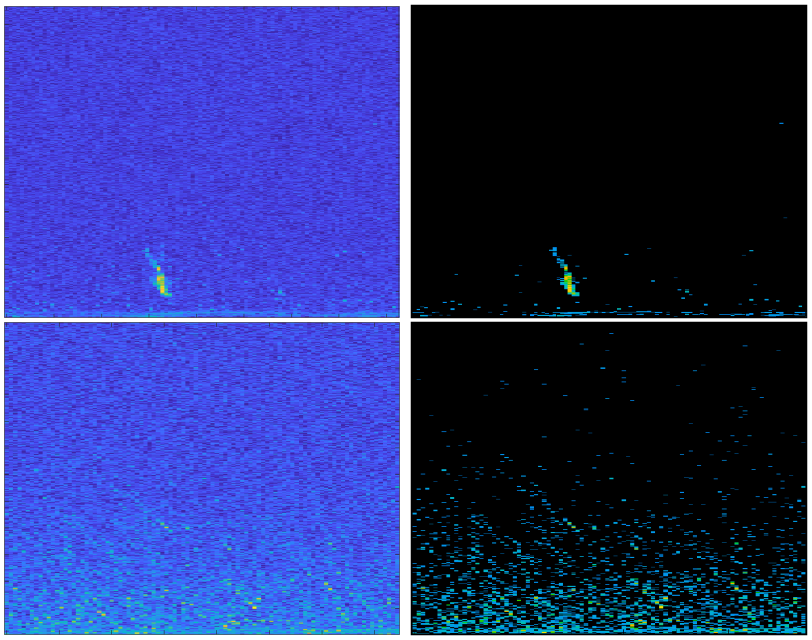

Figure 3.5: On the left column: raw images of echograms. On the right: processed images with k-means clustering algorithm

3. Each pixel color intensity in the cluster is subtracted to the mean value.

4. The previous steps are repeated until the sum of differences between pixels intensity and the mean will remain the same.

5. The cluster with more concentrated information is chosen. For that, we search for the cluster with higher frequency of pixels equal to zero.

6. To the chosen cluster, steps $1 - 4$ are performed. The final cluster is chosen based on its variance. Images with targets will have higher variance than background information, that tend to be more monochromatic.

The result of the pre-processing and image segmentation step can be observed in Fig. 3.6. We can first observe the raw image, on the left. On the right side we can observe that this process was able to separate the background, with more spread information with darker blue color in cluster 1 and the last image, the second cluster, that contains only the pixels with higher intensities (represented by yellow and light green pixels).

3.2.2 Obtaining the main features

Once the image contained only the information of interest, we set to identify the main features relevant information that remained. Using the information from the training dataset we identified the most important features as the brightest areas in the images, i.e. the highest intensities of echos, as follows:

- Whiskers
- Single prey
- Schooling prey

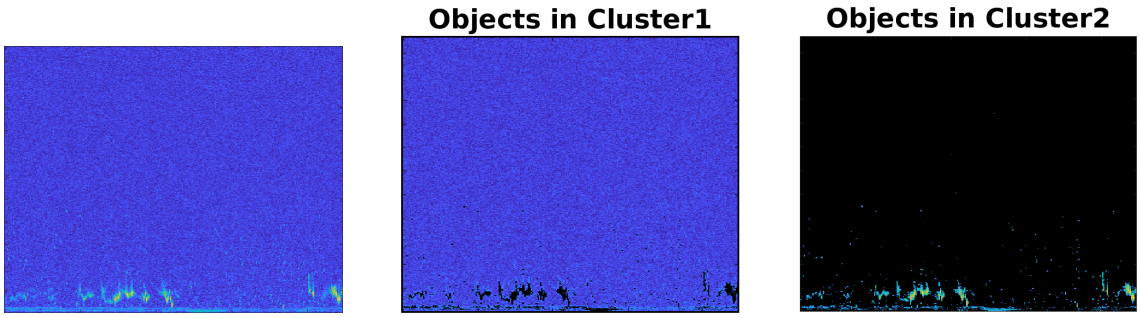

Figure 3.6: Raw image of an echogram and the processed image after two clustering processes. The final image that will be analyzed after that is cluster 2.

Figure 3.7: Main features in sonar images: Whiskers, a single prey and schooling prey.

The whiskers were the first feature we worked on. They are located near the bottom of the image -considering the bottom of the image represent the area close to the sonar transducer. The feature would be considered as whiskers if its position were on the bottom 95% of the image as it is shown in the first image in Fig. 3.7, the whiskers are a fine strip with colors close to light blue. The whiskers help the seal detecting preys and can be deployed right before and during a prey capture attempt. Because of that, sometimes it would appear connected to the prey, or any other target, which made difficult to separate it from the other features and possibly cause the mislabeling of the image. Therefore we calculated their mean location and pixel color and removed the pixels that matched that criteria and we decided to analyze the presence/absence of whiskers in separated images.

Once the whiskers were removed, we used isolated connected components, i. e. groups of adjacent pixels. From analyzing previously characterized images we established a threshold for size (area of at least 250 pixels) or color intensity (above 75% of maximum intensity). Visually we can observe that there is marked difference between isolated and schooling prey. Isolated prey have smaller and continuous traces, usually with stronger intensity, whereas schooling prey appear as a lot larger structure, with presence of holes inside the structure. Prey appear in various sizes, from 100 pixels for the smaller isolated prey, to 20000 pixels when there where schooling prey. So it was necessary to combine color and size information and have more flexible thresholds to separate the final features to be characterized. Some examples are shown in Fig. 3.7.

3.3 Image classification algorithm

After removing the whiskers the remaining features were summarized to a group of image attributes:

- Maximum area of bigger connected components, i.e area of continuous pixel distribution
- Mean pixel value of the channels Red, Green and Blue for these bigger connected components
- Pixel intensity in Gray scale and in decibels (sound intensity values)

After we created a table with spreadsheet of all 4976 echograms already characterized plus their main attributes presented above. A summary of table can be seen in the annex. From then on the data analysis was performed in python, using the *scikit learn*. Our first step was to scatter plot each data point in pairs to better visualize if they are correlated and if we could separate the data dispersion into the previously established labels for number of prey: one isolated prey, two isolated preys, or three or more prey (i. e. schooling prey), presented as classes 1, 2 and 3. In Fig. 3.8, we can visualize patches in the distribution of data, so we can conclude that this separation of data into groups is possible. To perform the classification we tested the K nearest neighbors (KNN) and Support Vector Machines (SVM) methods.

KNN is a supervised predictive model that uses the concept of similarities between samples and classifies each data sample based on the label from the k other data points that are close (*Triguero et al.*, 2019). The image is converted in a vector of features and the Euclidean distance (d) is calculated between image vectors in order to find the closest neighbors:

$$d(x, y) = (\sum_{i=1}^{m}(x_i - x_j)^2)^{\frac{1}{2}}, \tag{3.2}$$

where x_i is the image represented in a feature vector and x_j is the vector of already classified data. The neighbors will be the k closest data vectors. SVM is a supervised method where the distance between classes is maximized. The distance between classes is calculated between their closest points. This method uses the statistical learning theory to classify the images. The main objective of this method is to separate the distribution of attributes into classes that can be separated by linear boundaries. For that it uses the hyperplane, a linear subspace whose dimension will depend on the number of image attributes, the input sample features, we chose the color of pixels on three channels (RGB), sound intensity in decibels and target maximum size. Then the hyperplane that separates this $n = 5$ dimensional space has $n - 1 = 4$ dimensions. The classification is given by:

$$\begin{cases} y_i(x^T + _0) \geq 1 - \xi_i \forall i, \\ \xi_i \geq 0, \Sigma\xi_i \leq constant \end{cases} \tag{3.3}$$

where y_i is the class, x_i the attributes, $x^T + _0$ is the equation that describes the hyperplane, ξ_i is the proportion data on the wrong side of the margin. The objective is to get $\Sigma\xi_i$ to the lowest value. For the SVM we tested different functions for the boundary, a linear kernel function, and two non linear kernel functions, a 3^{rd} degree polynomial and radial basis function (rbf).

The assessment of the classification methods was realized using a confusion matrix, where we can see if and how many labels have been assembled correctly. It is presented as a 3X3 matrix where rows are the true classes of the images and the columns are the predicted classes by the classification models. The first cell has the true positives, the images correctly labelled as class 1. The cells of the main diagonal, apart from the first are the images correctly labeled as the other two classes. Cells two and three in row one

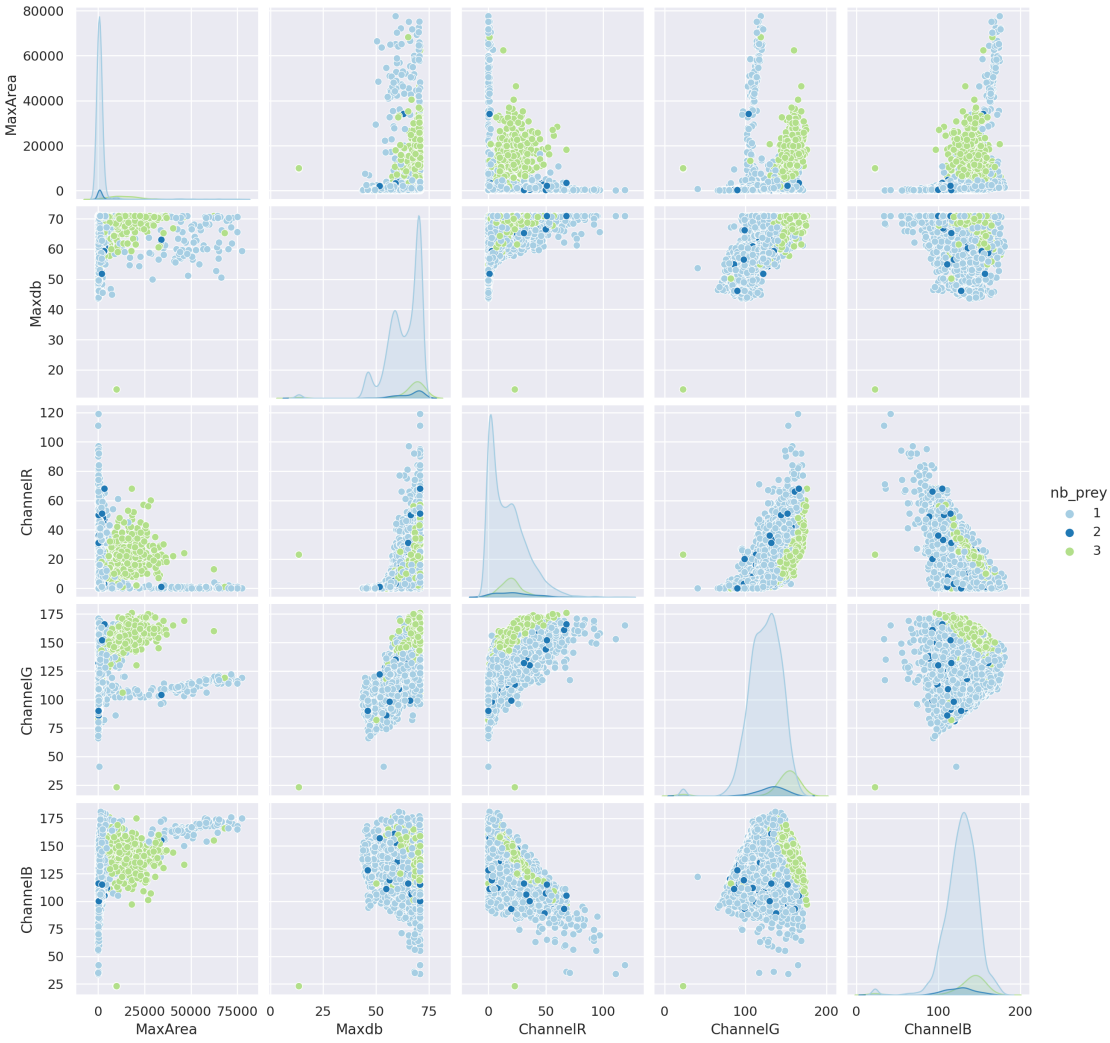

Figure 3.8: Scatterplot of main image attributes plotted in pairs. In color, labels representing the number of preys, 1 = *isolatedprey*, 2 = *morethanoneprey* and 3 = *schoolingprey*

are the false negatives, where the image was wrongly classified with labels 2 and 3. After that, we calculated the precision (the proportion between true positives and the total) and the recall (the number of positives the model got related to the total of positives in the training dataset) as evaluation metrics.

Chapter 4

Results

4.1 Evaluation of the image processing method

In order to assess the effectiveness of our image segmentation process we presented a sample of 1000 echograms from the processed images to two experts that contributed to the initial training dataset (500 images) to see if they would interpret visually the same way they did using the sonar game toolbox. A third expert that did not take part on the construction of training dataset, but followed the process of image treatment also analyzed 500 images, overlapping some of the echograms provided to the 2 initial experts.

We compare the performance of the the experts that made the training dataset with their previews answers in the table below. From the 1000 images that were reanalyzed, 24.4% of them were considered to contain no relevant information prey-wise anymore or didn't know the answer. From the remaining 76.6% there were 649 true positives, i. e. images that were correctly labelled as class 1, and 67 true negatives, i.e. the images were correctly labeled as classes 2 and 3. The labeling of Class 2 had the worst results, with only 5 images labeled correctly. The overall accuracy was 93%. The recall for classes 1, 2 and 3 was was 0.98, 0.13 and 0.90, respectively and the precision was 94%, 36% and 97%.

	Predicted Class 1	Predicted Class 2	Predicted Class 3
Actual Class 1	649	8	2
Actual Class 2	34	5	0
Actual Class 3	6	1	61

Table 4.1: Confusion matrix for the classification by experts using processed images compared to the training dataset they created using the sonargame toolbox

The independent expert had similar results for class 1, but worse precision an recall for the other two classes. From 500 images, 28% were considered as "empty" or they did not know the answer. There were 288 true positives and 27 true negatives. Again, class 2 was the less correctly assessed, with only one match with the training dataset. The overall accuracy was 88%. The precision values for classes 1, 2 and 3 were 95%, 0.04 and 88%, respectively. The recall values were 0.91, 0.08 and 0.88, respectively.

	Predicted Class 1	Predicted Class 2	Predicted Class 3
Actual Class 1	288	24	3
Actual Class 2	10	1	1
Actual Class 3	4	0	28

Table 4.2: Confusion matrix for the classification by the independent using processed images compared to the training dataset

4.2 Image classification with training dataset

All confusion matrices for the classification methods are presented below:

	Predicted Class 1	Predicted Class 2	Predicted Class 3
Actual Class 1	1292	2	20
Actual Class 2	58	0	2
Actual Class 3	18	0	102

Table 4.3: Confusion matrix of the KNN classification method

	Predicted Class 1	Predicted Class 2	Predicted Class 3
Actual Class 1	1290	0	24
Actual Class 2	60	0	2
Actual Class 3	31	0	89

Table 4.4: Confusion matrix of the SVM classification method with kernel=linear

	Predicted Class 1	Predicted Class 2	Predicted Class 3
Actual Class 1	1308	0	6
Actual Class 2	54	0	1
Actual Class 3	31	0	66

Table 4.5: Confusion matrix of the SVM classification method with kernel=3^{rd} degree polynomial

	Predicted Class 1	Predicted Class 2	Predicted Class 3
Actual Class 1	1290	0	24
Actual Class 2	58	0	2
Actual Class 3	14	0	106

Table 4.6: Confusion matrix of the SVM classification method with kernel= Radial basis function

We can observe that both methods had similar number of true positives, with SVM with linear kernel had a slightly better performance (1308 true positives). Both also had similar number of true negatives, meaning that they assigned correctly classes 2 and 3 similarly (108 for SVM with rbf kernel and 104 for KNN method). The most challenging situation was classifying the images as class 2. All methods assigned class 2 images mostly as class 1 and less by class 3.

The evaluation of the classification methods performance was through precision and recall. The results are shown in the table below. The classification methods had similar results for class 1, with precision higher equal or higher than 0.98. All methods had inferior performance within class 3, with precision ranging from 55% for SVM with a polynomial kernel to KNN, with precision of 85%. As for class 2, as none of the methods were able to predict it, they assembled class two images as class 1. The recall values for all methods were similar, diverging in class 3, with the range of 0.78 for SVM with linear kernel to SVM with polynomial kernel. Overall both methods, KNN and SVM, had good results, especially with class 1 images, i. e. images with a single prey.

	Precision			Recall		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
KNN	0.98	0	0.85	0.94	0	0.82
SVM (linear)	0.98	0	0.74	0.93	0	0.78
SVM (poly 3)	0.99	0	0.55	0.92	0	0.94
SVM (rbf)	0.98	0	0.74	0.93	0	0.74

Table 4.7: Performance of the classification methdods K nearest neighbors and Support Vector Machines (with different kernel functions)

4.2.1 Presence or absence of whiskers

When elephant seals are foraging, they use their highly sensitive vibrissae (the whiskers) and their vision to detect close prey movement (Lev). In echograms, they normally appear before possible prey traces close to the transducer, at the bottom of the image (Fig. 4.1. However, because the beam from the sonar tag is very narrow (3.4ř), the vibrissae presence on echograms depends on the device's position on the head of the seal. For example, seals KER2, KER3 and PV1 had fewer images with the presence of whiskers (18%, 1% and 12% respectively), compared to seals KER1 and PV2 (57% and 96%, respectively).

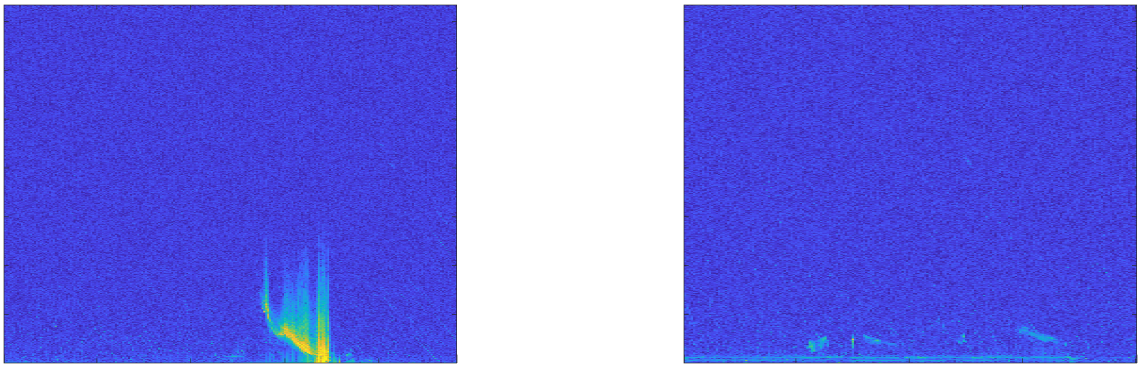

Figure 4.1: On the left, raw echogram where the whiskers deployment is not observed from seal KER3, on the right the deployment of whiskers is observed from seal PV2.

Chapter 5

Discussion

In this work we analyzed over 4976 echograms from microsonars tagged in 5 Southern elephant seals. These echograms were part of a training dataset to test the classification methods. We evaluated both the image processing and segmentation, comparing the performance of experts that did visual classification with processed images and unprocessed ones, and the chosen classification methods.

Our first objective was to isolate relevant signals related to predator-target encounters. For that, we needed to isolate three main image features: whiskers, isolated prey and schooling prey. Although the image processing was able, in general, to keep relevant information, its steps still needs fine tuning. When asked about their impressions on the processed images the experts found harder to differentiate ambient interference from weaker echos that may be targets. They guide their decisions mainly on color information, so the context in each image is very important and might be lost when it is over processed. They also found the black background misleading because it makes the weaker traces stand out, leading to error in the echogram interpretation. Due to the narrowness of the solar beam (3.4 deg), only targets that are directly in front of the seal will be detected and the target strength will be higher if the prey is swimming right in front of the seal or in a oblique angle (*Simmonds and MacLennan*, 2008). Therefore, the visual interpretation of echograms is a complex task, with uncertainty, so regularly information from more sensors is combined to obtain a final answer. When assessing prey type in this sonar data *Goulet et al.* (2020) used concurrent accelerometer data as complement information to check if the seal was making a strike at the time they detected a possible target, and have more clues over inferring the presence of prey. In order to keep weaker traces in some images and avoid empty images, after the assessment of the experts all images were reprocessed with lower thresholds .

Although our segmentation process may have made difficult the visual interpretation of the images, the main goal of this work was to start to develop a method based in machine learning to interpret them. So we investigated if classification algorithms (KNN and SVM) could interpret the extracted features and classify them into prey type (single vs schooling prey), they reactivity to the seal (passive versus active), and categories of acoustic signal size and intensity. We were able to classify the images into prey type, but not their reactivity or categories of acoustic size. This is due to the choice we made to reduce the main components in the image to their mean values, when it refers to colors, or the maximum value of sound intensity in decibels. To study their reactivity we would need to use more information from the main connected components, as their orientation,

for example. The orientation would give us an inference of their relative distance from the seal. Positive orientation means the prey is getting away from the seal, recalling that the vertical axis in echograms is the distance from the transducer.

Our third objective was to detect in the echogram whiskers deployment events. Our segmentation method was able to isolate well these structures, as we can see in the first image in Fig. 3.7. Our method profited from the predictability of their location, as well as their color and sound intensity that are within a relatively fixed range. The analysis of connect components in the image is of great value for this type of work, however a restriction is the fact that horizontal gaps would separate an otherwise large component. When this happened, we would account the presence of whiskers in an image based on a minimum size of an horizontally oriented connected component.

5.1 Limitations of the method and future perspectives

We investigated if we could correctly classify the echograms using only image attributes even if their visual interpretation were not ideal. Our objective was to create an automated classification process, and although the classification method had satisfactory results, more work should be implied into connecting the color range to echo intensity in order to avoid highlighting the wrong targets. In addition, using the color information and size of targets alone is not sufficient to fully characterize the echograms. As mentioned previously, echograms have complex interpretation, therefore an advancement would be add information of the contrast of stronger targets to the background noise, to avoid highlighting weak signals, and to use the accelerometer data to confirm if the seal was actively chasing any prey concomitantly with the sonar information. In addition, this knowledge could assist the distinction between echograms containing a prey and "empty" echograms, that is somewhat difficult to be made.

The large difference we observe in precision and recall for class 1 compared to classes 2 and 3 may be due to the fact that the training dataset is highly unbalanced, with 87% of images in class 1, 8% in class 3 and 4% in class 2. *Goulet* (2020) analyzed part of this sonar dataset and found that for the PrCAs, the seals encountered schooling prey (class 3 of prey type) only 3% of the times, the only exception was seal (PV2) with 49% of PrCAs analyzed with schooling prey. Therefore, it was not possible to have similar number of echograms with isolated and schooling prey. As we can observe from Fig. 3.8 there is no distinction in the distribution of image attributes of class 1 and class 2 images. Moreover, as the training dataset is unbalanced, the model will tend to choose the most abundant class. This method as it is can only differentiate a single isolated prey from schooling prey. Another drawback of our method is that as all echograms in the training dataset had at least one isolated prey, we could not train our classifiers with "empty" images. When applied to the larger dataset, a large number of echograms would be wrongly assigned to one of the three classes.

Both classification methods, K nearest neighbors and Support Vector Machines, had overall good results, but KNN method had better precision for class 3. Supervised methods are highly dependant on the training dataset, which was highly unbalanced. KNN methods privileges the class from nearest data samples, being biased towards class 1, that was the most abundant. Whereas the best performance of SVM methods was the linear

and rbf kernel possibly because they are better suited for acoustic signals and polynomial kernels are generally used for normalized data (*Cervantes et al.*, 2020). A disavantage of SVM method the higher processing time, compared to KNN, that took less than a second to train and test all images. As there is no gain in terms of performance, KNN is the best classification method for these echograms. Although these methods had good results for classes 1 and 3, they were not able to properly train the model with the distribution of class 2 attributes, so it will be necessary to make adjustments on the training dataset as to increase the number of images of class 2. Possibly the image information solely won't suffice to differentiate an isolated prey that appear multiple times in an echograms from more than one isolated prey, possibly tracking the main objects on the image.

This work is a initial step into using machine learning methods to analyze this sonar dataset. We managed successfully to detect and isolate structures related to the deployment of whiskers, the horizontal strip in the bottom of the image, as well as automate an algorithm to detect and count the number of times they appear per seal. In the future it is necessary to use the connected components as objects and have information of all pixels, not only the mean value. Another possibility is to use a deep learning method, such as neural networks, to automatically extract the features, classify and detect the targets and compare its performance with the supervised machine learning methods we used in this work. Then one can study if its more advantageous for sonar images to use supervised learning, albeit with an unbalanced training dataset, or to let the machine extract the features and create the labels for the Ṫo differentiate the echogram with one or more isolated prey, these detected objects could be tracked, using a Kalman Filter to study its trajectory, for example (*Chan et al.*, 1979). The next steps in this work should be:

- Redo the training dataset, in order to get a more balanced one. Include "empty" echograms and make it 4 classes, 0, 1, 2 and 3.
- Use color information, intensity and size from all connected components
- Once is detected a prey in the echogram, study the orientation of biggest targets to infer prey reactivity. If the prey is escaping, the trace will have positive orientation, moving away from the seal
- Apply this method in sonar images in other species that were tagged with the same device, such as fur seal and assess its applicability
- Use the full classified dataset and compare it to dive records, oceanographic data to better study foraging behaviour of elephant seals.

Bibliography

().

Benoit-Bird, K. J., and W. W. L. Au (2003), Prey dynamics affect foraging by a pelagic predator (stenella longirostris) over a range of spatial and temporal scales, *Behavioral Ecology and Sociobiology*, *53*(6), 364–373.

Block, B. A. (2005), Physiological Ecology in the 21st Century: Advancements in Biologging Science1, *Integrative and Comparative Biology*, *45*(2), 305–320, doi:10.1093/icb/45.2.305.

Campagna, C., A. R. Piola, M. Rosa Marin, M. Lewis, and T. Fernández (2006), Southern elephant seal trajectories, fronts and eddies in the brazil/malvinas confluence, *Deep-Sea Research Part I*, *53*(12), 1907–1924, doi:10.1016/j.dsr.2006.08.015.

Cervantes, J., F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez (2020), A comprehensive survey on support vector machine classification: Applications, challenges and trends, *Neurocomputing*, *408*, 189–215, doi:https://doi.org/10.1016/j.neucom.2019.10.118.

Chan, Y., A. Hu, and J. Plant (1979), A kalman filter based tracking scheme with input estimation, *IEEE Transactions on Aerospace and Electronic Systems*, *AES-15*(2), 237–244, doi:10.1109/TAES.1979.308710.

Chandra, M. A., and S. Bedi (2018), Survey on svm and their application in image classification, *International Journal of Information Technology*, pp. 1–11.

Dhaware, C., and K. Wanjale (2016), Survey on image classification methods in image processing, *Int. J. Comput. Sci. Trends Technol*, *4*(3), 246–248.

Dillencourt, M. B., H. Samet, and M. Tamminen (1992), A general approach to connected-component labeling for arbitrary image representations, *Journal of the ACM (JACM)*, *39*(2), 253–280.

Fedak, M. A., P. Lovell, and S. M. Grant (2001), Two approaches to compressing and interpreting time-depth information as collected by time-depth recorders and satellite-linked dat recorders, *Marine Mammal Science*, *17*(1), 94–110, doi:https://doi.org/10.1111/j.1748-7692.2001.tb00982.x.

Field, I., C. Bradshaw, H. Burton, and M. Hindell (2004), Seasonal use of oceanographic and fisheries management zones by juvenile southern elephant seals (mirounga leonina) from macquarie island, *Polar Biology*, *27*(7), 432–440, doi:10.1007/s00300-004-0615-3.

Gille, S. T. (2002), Warming of the southern ocean since the 1950s, *Science*, *295*(5558), 1275–1277, doi:10.1126/science.1065863.

Goulet, P. (2020), Dare to differ: Individual foraging success in marine predators, from the largest seal to small penguins, Ph.D. thesis, University of St Andrews.

Goulet, P., C. Guinet, R. Swift, P. T. Madsen, and M. Johnson (2019), A miniature biomimetic sonar and movement tag to study the biotic environment and predator-prey interactions in aquatic animals, *Deep Sea Research Part I: Oceanographic Research Papers*, *148*, 1–11, doi:https://doi.org/10.1016/j.dsr.2019.04.007.

Goulet, P., C. Guinet, C. Campagna, J. Campagna, P. L. Tyack, and M. Johnson (2020), Flash and grab: deep-diving southern elephant seals trigger anti-predator flashes in bioluminescent prey, *Journal of Experimental Biology*, *223*(10), doi:10.1242/jeb.222810, jeb222810.

Guinet, C., Y. Cherel, V. Ridoux, and P. Jouventin (1996), Consumption of marine resources by seabirds and seals in crozet and kerguelen waters: changes in relation to consumer biomass 1962–85, *Antarctic Science*, *8*(1), 23–30, doi:10.1017/S0954102096000053.

Hassaballah, M., and A. I. Awad (2016), *Detection and Description of Image Features: An Introduction*, pp. 1–8, Springer International Publishing, Cham, doi:10.1007/978-3-319-28854-3_1.

Hindell, M., D. Slip, and H. Burton (1991), The diving behavior of adult male and female southern elephant seals, mirounga-leonina (pinnipedia, phocidae), *Australian Journal of Zoology*, *39*(5), 595–619.

Kastak, D., and R. J. Schusterman (1999), In-air and underwater hearing sensitivity of a northern elephant seal (mirounga angustirostris), *Canadian Journal of Zoology*, *77*(11), 1751–1758, doi:10.1139/z99-151.

Kim[1], J., B. Kim, and S. Savarese (2012), Comparing image classification methods: K-nearest-neighbor and support-vector-machines, in *Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics*, vol. 1001, pp. 48,109–2122.

Kotsiantis, S. B., I. Zaharakis, P. Pintelas, et al. (2007), Supervised machine learning: A review of classification techniques, *Emerging artificial intelligence applications in computer engineering*, *160*(1), 3–24.

McConnell, B. J., C. Chambers, and M. A. Fedak (1992), Foraging ecology of southern elephant seals in relation to the bathymetry and productivity of the southern ocean, *Antarctic Science*, *4*(4), 393–398, doi:10.1017/S0954102092000580.

McIntyre, T., P. De Bruyn, I. Ansorge, M. Bester, H. Bornemann, J. Plötz, and C. Tosh (2010), A lifetime at depth: vertical distribution of southern elephant seals in the water column, *Polar Biology*, *33*(8), 1037–1048.

McMahon, C., I. Field, C. Bradshaw, G. WHITE, and M. Hindell (2008), Tracking and data-logging devices attached to elephant seals do not affect individual mass gain or survival, *Journal of Experimental Marine Biology and Ecology*, *360*(2), 71–77.

Na, S., L. Xumin, and G. Yong (2010), Research on k-means clustering algorithm: An improved k-means clustering algorithm, in *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, pp. 63–67, doi:10.1109/IITSI.2010.74.

Otsu, N. (1979), A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66, doi:10.1109/TSMC.1979.4310076.

Pasolli, E., F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery (2013), Svm active learning approach for image classification using spatial information, *IEEE Transactions on Geoscience and Remote Sensing*, *52*(4), 2217–2233.

Petiteau, L. (2020), Performances analysis of the µsonar. an echo-sounding bio-logger.

Portner, H.-O., et al. (2019), Polar region, in *Special Report on the Ocean and Cryosphere in a Changing Climate*, Intergovernmental Panel on Climate Change, Geneva.

Ropert-Coudert, Y., M. Beaulieu, N. Hanuise, and A. Kato (2010), Diving into the world of biologging, *Endangered Species Research*, *10*, 21–27, doi:10.3354/esr00188.

Simmonds, J., and D. N. MacLennan (2008), *Fisheries acoustics: theory and practice*, John Wiley & Sons.

St. John, M. A., A. Borja, G. Chust, M. Heath, I. Grigorov, P. Mariani, A. P. Martin, and R. S. Santos (2016), A dark hole in our understanding of marine ecosystems and their services: Perspectives from the mesopelagic community, *Frontiers in Marine Science*, *3*, 31, doi:10.3389/fmars.2016.00031.

Triguero, I., D. García-Gil, J. Maillo, J. Luengo, S. García, and F. Herrera (2019), Transforming big data into smart data: An insight on the use of the k-nearest neighbors algorithm to obtain quality data, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(2), e1289.

Vacquié-Garcia, J., C. Guinet, A.-C. Dragon, M. Viviant, N. El Ksabi, and F. Bailleul (2015), Predicting prey capture rates of southern elephant seals from track and dive parameters, *Marine Ecology Progress Series*, *541*, 265–277, doi:10.3354/meps11511.

Wisniewska, D. M., M. Johnson, J. Teilmann, L. Rojano-Doñate, J. Shearer, S. Sveegaard, L. A. Miller, U. Siebert, and P. Madsen (2016), Ultra-high foraging rates of harbor porpoises make them vulnerable to anthropogenic disturbance, *Current Biology*, *26*(11), 1441–1446, doi:https://doi.org/10.1016/j.cub.2016.03.069.

Younger, J. L., L. M. Emmerson, and K. J. Miller (2016), The influence of historical climate changes on southern ocean marine predator populations: a comparative analysis, *Global Change Biology*, *22*(2), 474–493, doi:https://doi.org/10.1111/gcb.13104.