ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Animal Bioacoustics
Session 2aAB: Conditioning, Segmentation, and Feature Extraction in Bioacoustic Signals

2aAB6. Robustness of perceptual features used for automatic aural classification to propagation effects

Carolyn M. Binder*, Paul Hines, Sean P. Pecknold and Jeff Scrutton

*Corresponding author's address: Underwater Sensing, Defence R&D Canada, PO Box 1012, Dartmouth, B2Y 3Z7, NS, Canada, carolyn.binder@drdc-rddc.gc.ca

Previous effort has shown that a prototype aural classifier developed at Defence R&D Canada can be used to reduce false alarm rates and successfully discriminate cetacean vocalizations from several species. The aural classifier achieves accurate results by using perceptual signal features that model the features employed by the human auditory system. Current work focuses on determining the robustness of the perceptual features to propagation effects for two of the cetacean species studied previously - bowhead and humpback whales. To this end, classification results are compared for the original vocalizations to classification results obtained after the vocalizations were re-transmitted underwater over ranges of 2 to 10 km. Additional insight into the propagation effects is gained from transmission of synthetic bowhead and humpback vocalizations, designed to have features similar to the most important aural features for classification of bowhead and humpback vocalizations. Each perceptual feature is examined individually to determine its robustness to propagation effects compared to the other aural features. To gain further understanding of propagation effects on the features, preliminary propagation modelling results are presented in addition to experimental data.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Previous effort has shown that a prototype aural classifier developed at Defence R&D Canada [1] can be used to successfully discriminate cetacean vocalizations from several species; bowhead, humpback, North Atlantic right, and sperm whales vocalizations have been classified in this way [2]. To achieve these accurate results, the aural classifier employs perceptual signal features, which model the features used by the human auditory system. Like the human auditory system, the perceptual features used by the aural classifier can be employed to classify sounds from different sources, e.g. active sonar echoes [1] and cetacean vocalizations [2]. Thus, the aural classifier has proven to be very robust in its applicability to different data sources.

The current work aims to examine the robustness of the classifier under various environmental conditions, especially against propagation effects. To this end, an experiment was undertaken during the spring of 2012 aboard the Canadian Forces Auxiliary Vessel (CFAV) QUEST. During the experiment sets of real and synthetic bowhead and humpback vocalizations were transmitted under water over ranges of 2-10 km. These experimental data are used to develop a preliminary understanding of the robustness of the perceptual features to the effects of propagation. This is an important step in evaluating the classifier’s performance and robustness in an automated system that could be implemented onboard a ship or an autonomous vehicle.

In the following sections the data set used for the experiment will be described, followed by a description of the experimental method. Some preliminary results for the real bowhead and humpback vocalizations will then be discussed. The last section of this paper proposes future work that is intended to augment the preliminary results discussed here. The future work will include further experimental measurements as well as a proposed modeling effort.

DATASET

Real Bowhead and Humpback Vocalizations

The bowhead and humpback vocalizations were obtained from the MobySound website [3]; the data obtained from MobySound contained examples of bowhead song endnotes and humpback songs. From the humpback song in the recordings, four song units were selected that were repeated frequently and were similar in frequency and duration to the bowhead vocalizations. Thus, the data set was formed of the bowhead song endnotes and the four selected humpback units. From these data, 155 of the highest SNR vocalizations from each species were selected for use in the propagation experiment. Since it was not possible to obtain localization information for each signal, it was assumed that high SNR indicates that the vocalizing whale was relatively close to the recording unit, so that fewer propagation effects were contained in the recorded signals.

Bowhead and humpback whales vocalize at relatively low frequencies. The bowhead and humpback vocalizations in this data set had frequency content predominantly in the 50-800 Hz and 100-2000 Hz range, respectively [4]. Since no projector was available that would transmit signals over the approximately five-octave band of the vocalizations, the signals were filtered and scaled for transmission from an ITC-2010 projector. This sound source has a flat frequency response across the 2 octaves from 1-4 kHz. Pre-processing of the signals was done to take advantage of this 2-octave flat frequency band when transmitting the signals. The power spectra of each species’ vocalizations were RMS averaged (Figure 1) to identify a two-octave band that contained sufficient signal information to be representative of the full bandwidth calls. The energy was integrated for the full-bandwidth signals and compared with the energy in the 200-800 Hz range; this reduced frequency band contained 74% of the energy for the bowhead vocalizations and 72% for the humpback vocalizations. Thus, a significant proportion of the signal energy was contained in the 200-800 Hz band.

To test if this reduced frequency band contained sufficient information for aural classification, each of the vocalizations were bandpass filtered and classification was performed on both the full-bandwidth and reduced-bandwidth signals; the results are shown in Figure 2. Correct classification is represented by a coloured bar plotted on the background of the corresponding colour (eg. dark blue on light blue); misclassification is represented when a coloured bar is plotted on a region of a different colour (eg. blue on pink background). Two methods are used to quantify the performance of the classifier – accuracy and area under the Receiver Operating Characteristic (ROC) curve (AUC). The AUC provides a single quantitative performance metric that can be interpreted as the probability that the classifier will correctly classify a randomly selected pair of bowhead and humpback vocalizations.
FIGURE 1. RMS averaged power spectra for (a) bowhead and (b) humpback vocalizations. The black lines represent the average power spectra of the full bandwidth signals, and the blue lines are the power spectra after band-pass filtering the signals between 200-800 Hz.

The \( AUC \) varies between 1.00 (indicative of an ideal classifier) and 0.50 (equivalent to randomly assigning a classification decision). In general, a larger \( AUC \) implies better average classifier performance [5]. Full-bandwidth signals resulted in 94% classification accuracy and an \( AUC = 0.99 \); when using the reduced-bandwidth signals the classification performance reduced slightly, to 92% accuracy \( AUC = 0.98 \). Many of the same perceptual features were highly ranked discriminators for the full bandwidth and reduced bandwidth vocalizations. This is important because it suggests that applying a 200-800 Hz band-pass filter to the vocalizations does not remove the information required for calculating the important perceptual features. Altogether, this evidence suggests that enough of the signal content was contained in the 200-800 Hz range to accurately represent both species’ vocalizations. The final step in processing the signals for transmission was to increase the playback speed of each of the filtered signals by a factor of five to shift the signals into the pass-band of the ITC-2010 source. Of course, one would prefer to transmit the vocalizations in their original frequency band; however, scaling up in frequency does have the advantage that any propagation effects will be evident at shorter ranges, which simplifies the experimental geometry somewhat.

FIGURE 2. Classification results for (a) full bandwidth signals and (b) band-pass filtered (200-800 Hz) signals.
Synthetic Vocalizations

The bowhead and humpback vocalizations used in this experiment were already subject to unknown propagation effects when they were recorded. Therefore, in order to gain additional insight into the propagation effects, synthetic vocalizations were also used. In contrast to the real vocalizations, synthetic vocalizations provided a known signal with no propagation effects. The synthetic signals were designed to have similar mean and variance for three of the perceptual features that were considered important in discriminating bowhead and humpback vocalizations.

A total of 155 synthetic bowhead signals were created by concatenating four linear frequency modulated (LFM) signals and 155 synthetic humpback signals were created using LFM, exponential frequency modulated (EFM), and hyperbolic frequency modulated (HFM) signals. The length, and maximum and minimum frequencies of each frequency-modulated section were randomly varied to create feature-value distributions similar to those calculated from the real bowhead and humpback vocalizations. Figure 3 shows examples of both the real and synthetic bowhead and humpback vocalizations.

![Figure 3](image-url)

**FIGURE 3.** Spectrogram of example band-pass filtered signals used for propagation experiments. These are (a) real bowhead, (b) real humpback, (c) synthetic bowhead, and (d) synthetic humpback vocalizations.

EXPERIMENTAL METHOD

Both the real and synthetic signals were transmitted underwater from an ITC-2010 source deployed from the quarterdeck of CFAV QUEST as the ship drifted. The signals were received on a combination of moored and free-floating recorders that were positioned 2-10 km away from the ship. A depiction of the experimental configuration is shown in Figure 4. Only the data from the moored recorders are presented in this paper.

The experiment was performed twice, each time on a different day and at a different location. Both experimental sites were on the Scotian Shelf in an area known as Emerald Basin. The coordinates of the May 28 site are 43°35′N 63°05′W and for the June 2 site are 43°38′N 62°54′W. The water, source, and receiver depths at these sites are given in Table 1, as is the approximate distance over which the signals were transmitted.

An energy detector was implemented to detect the received signals and spectrograms of each of the detections were examined to verify that the detections corresponded to the transmitted signals. Each transmitted signal was then placed in its own .wav file for further processing. The sampling rate of the audio files containing the detections was then reduced by a factor of five so that the signals were scaled back to the original 200-800 Hz frequency band.

The aural classification process begins with calculating the perceptual features; to do this, an auditory model is applied to obtain a perceptual representation of each of the signals (details of the model in Ref. [1]). The classifier is trained with the original signals, for which the classifier is provided the class label. The classifier is then tested on the transmitted signals for which the classifier has no direct knowledge of the associated class labels. The remaining steps are performed using the training set, the results of which are applied to the testing set. Discriminant analysis (DA) [6] is used on a subset of features that best discriminate between classes to reduce the dimensionality of the...
FIGURE 4. Representation of the experimental setup. Rx\textsubscript{Moored} refers to the moored recorders and Rx\textsubscript{SB} refers to free-floating recorders. The distances between the ship and moored recorders ($r_1$ and $r_2$) ranged between 2 and 10 km. The coloured background gives an example of the modeled transmission loss at the June 2 experimental site.

Analysis of the transmitted synthetic signals has not yet been completed, so the following discussion focuses on the real vocalizations. Five features – duration, global mean sub-band decay time, local maximum sub-band decay time, frequency of the global maximum sub-band attack time, and peak loudness value (see Ref. [1] for details on these features) – that best discriminated between the real bowhead and humpback vocalizations were identified from the training set. The classification results presented in the following section were obtained by using these five features.

TABLE 1. Approximate water, source and receiver depths, and distance from source to receiver at the experimental sites.

<table>
<thead>
<tr>
<th></th>
<th>May 28, 2012 (43°35′N 63°05′W)</th>
<th>June 2, 2012 (43°38′N 62°54′W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water Depth</td>
<td>234 m</td>
<td>225 m</td>
</tr>
<tr>
<td>Source Depth</td>
<td>34 m</td>
<td>34 m</td>
</tr>
<tr>
<td>Receiver 1 Depth</td>
<td>48 m</td>
<td>50 m</td>
</tr>
<tr>
<td>Receiver 2 Depth</td>
<td>68 m</td>
<td>N/A</td>
</tr>
<tr>
<td>Propagation Range</td>
<td>2.5 km</td>
<td>9.5 km</td>
</tr>
</tbody>
</table>

RESULTS AND DISCUSSION

The results obtained by training the classifier on the band-pass filtered bowhead and humpback vocalizations (i.e., the signals that were not transmitted through the water) and testing the classifier on the transmitted signals are shown in Figure 5. Classification results for the training set are shown in panel (a) – the classification resulted in an accuracy of 95%. The classification accuracy decreased from 73% when the signals were transmitted over 2.5 km to 58% when the propagation range increased to 9.5 km. The ROC curves for the training set and the transmitted signals are shown in panel (d) of Figure 5. The AUC values were 0.99 for the training case and 0.90 for both propagation ranges. The similarity of the AUC values for both propagation experiments suggests that the separation between the bowhead and humpback classes was maintained, although it has already been noted that there was significant difference in the accuracies. This can be seen in the decision regions in panels (b) and (c) – both figures exhibit a similar amount of overlap of the data in the classes, although the accuracy has decreased. This decrease in accuracy appears to result from a shifting of the classes with respect to the decision threshold (i.e. boundary between
FIGURE 5. Classification results for the real bowhead and humpback signals. (a) The classifier was trained using the original signals, then tested on the signals that were transmitted over a (b) 2.5 km and (c) 9.5 km range. The results shown in (b) are from the hydrophone located at 48 m depth. (d) ROC curves – the black curve corresponds to the training set shown in panel (a) and the red and blue curves correspond to results in panels (b) and (c), respectively.

decision regions), while approximately maintaining the shape of the class distributions and separation between classes. These preliminary results suggest that classification performance may be affected by propagation range. If this is the case, then the classifier should be trained with signals collected at both short and long ranges. Propagation effects on the performance of the aural classifier may be analogous to the dependence of the aural classifier’s performance on the Signal to Noise Ratio (SNR) levels of active sonar echoes. Work done by Murphy and Hines [7] demonstrated that higher performance was achieved when the classifier was trained with echoes of a similar SNR to the SNR of the data in a test set. It was also found that if echoes in the test set had a range of SNR values, the best performance was achieved by training the classifier with echoes that also encompassed a wide range of SNR values.

FUTURE WORK

The work discussed in this paper is a result of initial analysis from a limited experimental data set. Analysis of this data set will continue and will include examining classification results using the synthetic signals. Future work will also investigate the robustness of the classifier through examination of further experimental results as well as data produced by propagation modeling. These additional data sets should provide sufficient evidence to rank individual perceptual features based on their robustness against propagation effects.

Another propagation experiment, like the one described in this paper, is planned to take place in the Gulf of Mexico during the spring of 2013. The experiment will be very similar to the one described here – it will be repeated at a different site that is therefore subject to different environmental conditions. These follow-on experiments will encompass more/different propagation conditions and ranges than were represented in the preliminary data set. Additionally, there will be further development of the synthetic vocalizations, prior to the spring 2013 experiment, so that the synthetic vocalizations’ perceptual features will be more representative of the real bowhead and humpback vocalizations’ features.

Propagation models will play an important part in determining the robustness of the classifier to propagation effects, since it is easier (and more financially feasible) to adjust the parameters of a propagation model than to repeat the experiment many times at many different locations and under different propagation conditions. A pulse
propagation model – Ocean Acoustics and Seismic Exploration Synthesis (OASES) – will be used to model the propagation effects that are applied to a signal as it is transmitted through the water. The aural classifier will be used on these artificially transmitted signals to gain further understanding of the propagation effects on the perceptual features.

To validate the use of the pulse propagation model, the environment at the experimental sites will first be modeled from the environmental data (e.g. sound speed profile) that were collected. The signals used during the experiment will be transmitted through the modeled environment and the effect on the perceptual features will be compared to those from the experimental data. Once the pulse propagation model has been validated, the parameters may be adjusted to study a variety of transmission conditions.

CONCLUSIONS

A data set containing relatively high SNR bowhead and humpback vocalizations and synthetic vocalizations were transmitted over ranges of 2-10 km. A preliminary analysis of the data collected from this propagation experiment suggests that classification performance may depend on the range over which the signals are transmitted. Classification accuracy decreased as the propagation range increased, although the AUC values remained constant with propagation range. This suggests that the classifier should be trained using signals that were transmitted over varying ranges. Further work needs to be done to ascertain how robust the classifier is to propagation effects and if some of the perceptual signal features used by the aural classifier are more robust than others. A follow-on experiment is planned for the spring of 2013 to collect additional data under different environmental conditions and for more propagation ranges. In addition to the experimental data, results from a pulse propagation model will be used to further investigate the robustness of the aural classifier to propagation effects.

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge Chris Browne, Daniel Graham, Mark Fotheringham, Roger Arsenault, and Paul Anstey all of DRDC - Atlantic, as well as the officers and crew of CFAV QUEST who took part in the Q345 sea trial. The authors wish to thank the creators of MobySound for the availability of data. This work was supported in part by the U.S. Office of Naval Research, Code 32, Marine Mammals & Biological Oceanography Program under Grant N00014-12-1-0139.

REFERENCES