W. Randolph Warner
3513 Di Leuca St.
Punta Gorda, FL 33950

Ph: 941-505-0793
Fx: 941-505-0718
e-mail: onconex@sunline.net

Final Report

02/14/03

FWC 01/02-14
Manatee Avoidance Technology
Vendor #: F257760246001

Title: *Manatee Proximity Indicator*

Project Overview

Technology designed to inform boaters of the presence of manatees in order to enable boaters to avoid manatees without changing the behavior of the animals.

It's very well documented that Marine Mammals vocalize. Manatees, Whales and Dolphins. Manatee vocalizations are generally characterized by "Squeaks," "Squeals," "Chirps" "Whistles," etc.. By recognizing the unique vocalizations of the Manatee while in their underwater environment and using it to trigger a signal to alert their presence can help avoid injuries and deaths associated with boat collisions, being crushed in flood gates, canal locks, marine construction sites, dredging sites, marine sporting events, industrial / municipal water in-takes, navigational waterways and all other areas where if their presence were known, possible harm may be avoided.

Work Proposed

To gather the known Manatee Vocalizations (see Manatee Vocalization Chart) digitally recorded and program Voice Recognition Integrated Circuits (VRIC) to recognize these unique Manatee vocal acoustical commands. The programmed acoustical commands will relay and operate micro-circuitry that will deliver a predetermined out-put or signal when activated.

Develop the working hardware prototype of the VRIC's, micro-circuits and Hydrophone system.

Summary of Activities

Initial Central Processing Unit development, relay design conference with the design engineers and plans for prototype construction.

Prototype CPU, Transducer and signal housing design completed.

Housing internal humidity/exposure seal, corrosion factors and the methods and test design completed. The humidity control scale using moisture absorbent silica gel discantant cans in real time / real exposure, pulls at 15 days, 30 days and 60 days. The methodology of determining weights of control silica cans and housing cans to establish design tolerance and acceptable parameters for the housing seals and internal electronics. Two housing systems were constructed . One housing system was placed at +/.5m above sea level at RH (Relative

Humidity) 55%-100% containing 4 DC's (discantant cans). Control DC's were stored in a Lunaire Environmental series CEO910-3 stability cabinet @ +25° C/60% RH.
At each pull the active & control DC's were weighed using an Acculab AL-204 series analytical balance to find moisture gain/loss.
Hydrophone purchased. REASON TC-4013 received and field validated : sensitivity 5-100kHz /131.17dB re 1uPa/V @1m, amplitude 29.95, pulse width 1.40ms at an depth of 1.3m. The REASON TC-4013 was used in phase 1 of recordings.
DAT (digital audio tape) recorder purchased . SONY TCD-D100 received, calibrated and field validated. Digital peak level limitation was set @<87db and sampling frequency was established @ 48kHz . Recording volumes set 0dB >. play back response @ 48kHz 20-22,000Hz @ 1db. Harmonic distortion <0.0008% @ 1kHz, 22Hz w/ peak level limitation engaged.
Comparative harmonics to control analog signals validated the REASON TC-4013 hydrophones input had an independence 0f 4.7 kilohms, input level of 1.4mVPlanned vocalization data collection, phase 1 of captive animals completed. All sounds transferred digitally from Digital Audio Tape (DAT) to hard disk at 48 kHz via Universal serial Buss (USB) at nominal input. Noise reduction algorithm applied to individual samples to remove extraneous background noise. All sounds normalized to 98% to attain an increase in Signal to Noise ratio of 5dB to 20dB (noise is reduced 21dB and signal 1dB for example).All sounds individually processed to observe preservation of original digital recording. Resulting data transferred to Audio CD at 44.1kHz after down sampling from 48kHz to achieve CD quality audio. Data CD also processed in Microsoft wav format at 44.1 kHz to mirror audio CD.
Recording Trips: 3/23-29/02 and 5/2-3/02 completed. Recordings were collected at Homosassa Springs Wildlife State Park with permission from the Park Manager, Tom Linley. Field recordings were conducted with digital audio recorder and RESON hydrophone equipment (listed above). The hydrophone was suspended horizontally from a telescoping pole over the water approximately 1m above the surface at varying distances/lengths 1-5m dependant on landside positioning. All recordings were preformed from landside. Hydrophone depth 1-1.5m. This phase 1, captive recording trip collected 259 digital manatee vocalizations.
U.S. Fish &Wildlife Permit applied for and amendment completed and sent.
Digital re-mastering of vocalization provided by/ recorded by Dr. Tom O'Shea completed.
Digital formatting of 259 vocalizations collected w/captive animals completed.
60 vocalizations digitalized and cataloged for Integrated Circuit programming.
Copy of "Digital Re-Mastered demonstration compact disk delivered to Katherine Frisch.
Analogue prototype circuit board completed with "playback" demonstrated to Katherine Frisch on her site visit onAug.29.
Sensory Inc. contracted to program voice recognition circuit.

<u>Problems Encountered</u>
Programming budget over-run of $20,500.00

Actual "field recordings needed" as reported in several project Interim Reports "6/17/02 delay in field recording of animals in the "wild" due to need of U.S. Fish &Wildlife Permit." and "8/15/02 delay in field recording of animals in the "wild" due to need of U.S. Fish &Wildlife Permit." No further vocalizations recorded due to U.S. F&W permit was not issued. A greater sampling of vocalizations in the wild and a wider scope of age and gender of animals are needed prior to the expense of having the Integrated Circuit programmed with a limited sample of voiceprints.

No permit was issued and the decision to proceed with a "Limited Scope Device". Due to the lack of the U.S. F&W permit to record vocalizations in the wild with natural background sounds, Sensory Inc. was given the contract to proceed to program a "limited recognition" device.

Bench Test

A bench test was conducted at the University of South Florida Marine Science Center on December 13, 2002. Dr. Forrest Mozer was designated by Sensory as Lead Engineer on the project. Dr. David Mann of USF and Katherine Frisch preformed the testing.

The Manatee Proximity Indicator (MPI) is a self-contained prototype without any external support for computers, laptops or software programs.

All test manatee vocalizations programmed for playback were recognized by the MPI, with the exception of a low volume, low frequency "rusty pump" type of vocalization, which was not detected.

Due to the limiting factors of not having actual "field recordings" for the MPI programming and playback test, the following conclusions were reported by Dr. Mozer:

The "Quiet Background Noise" (quiet_bn.wav) and "Loud Background Noise" (loud_bn.wav) provide me with a good understanding of the kinds of background sounds that the recognizer has to expect.

The "Manatee sounds alone" (manatee_test.wav) seem clearly to be manatee vocalizations removed from their noise backgrounds and strung together. So, during each vocalization, the audio signal contains the manatee vocalization plus whatever noise there was in that particular recording.

The "Quiet with manatee sounds" (quiet.wav) seems clearly to bethe manatee_test.wav recording of the previous paragraph, added to the quiet_bn.wav recording. And the "Loud with manatee sounds" (loud.wav), is the loud_bn.wav file added to manatee_test.wav.

The quiet.wav file and loud.wav file suffer two problems. First of all, there is double noise underneath all of the manatee vocalizations. One piece of noise comes from the original recording. The second piece of noise comes from the addition of either quiet.wav or loud.wav to the manatee vocalizations. Our algorithm tries to subtract the noise before the vocalization from the (noise+signal) that is occurs during the vocalization, and it can't do that because, in none of the three cases (manatee_test.wav, quiet.wav, or loud.wav), is the noise before the vocalization the same level as the noise during the vocalization. If we trained a neural net on these recordings, it might not work on actual manatee vocalizations in water or on the original recordings from which the manatee vocalizations in manatee_test.wav were extracted.

A second and potentially more serious problem is that the signal to noise ratio in quiet.wav (or loud.wav) depends on how the person who mixed quiet_bn.wav (or loud_bn.wav) with manatee_test.wav set the knobs on the mixer. THESE ARE NOT NATURAL RECORDINGS. The signal to noise ratio in these recordings may not be representative of that in raw recordings of manatee vocalizations.

The way we train the neural net depends sensitively on the signal to noise ratio and I cannot determine typical values of the signal to noise from this data.

All of these problems would be obviated if they only used natural recordings that were not touched up and did not have external noise added to them by a human. For all these reasons, my opinion is that it is not worth training a neural net based on guesswork and these

recordings, because it probably would not work in the real world environment. Is it possible to send me natural, untouched recordings and to use such recordings in evaluating our recognition performance?
Because the number of shrimp clicks is far greater than the number of manatee vocalizations, there may be a relatively large number of false recognitions (the algorithm deciding that a shrimp click or a group of shrimp clicks is a manatee). The problem of false triggers can be overcome by a strategy that depends on knowing a little about the frequency of manatee vocalizations. For example, if vocalizations are frequent when they are present, one could make up a rule like "If 5 or more positive recognitions are found within one minute, this is a manatee signal. If fewer than 5 are found, these are false triggers due to shrimp."
Of course my number of 5 in one minute is just made up and the real numbers would come from understanding how frequently one hears manatee utterances during a time when manatees are present. This will be a powerful technique for minimizing false triggers, which could otherwise be a major problem.

Conclusion

As demonstrated in the "bench-test", the "Proof of Concept" of the Manatee Proximity Indicator was not a Laptop driven software program, but an actual working self-contained prototype device that did distinguish the programmed manatee vocalizations and register a "signal" (via strobe light). However, being restricted in the ability to obtain natural environment recordings (No US F&W Permit) for programming & testing hampered the desired results as described in the Problems Encountered section.
The "MPI" is a working prototype and demonstrated valid recognition. The working hardware and circuitry of the MPI is developed. Further data collection and circuitry input/programming could prove be to a valuable adjunct in manatee avoidance technologies.

Overview for Project Continuance

This specification is written under the assumption of project continuance that original and additional unmodified field recordings of manatees in noise environments ranging from quiet to loud will be available for this project. It is also assumed that the recognizer (MPI) will be tested with either original, unmodified recordings or in actual use in water.

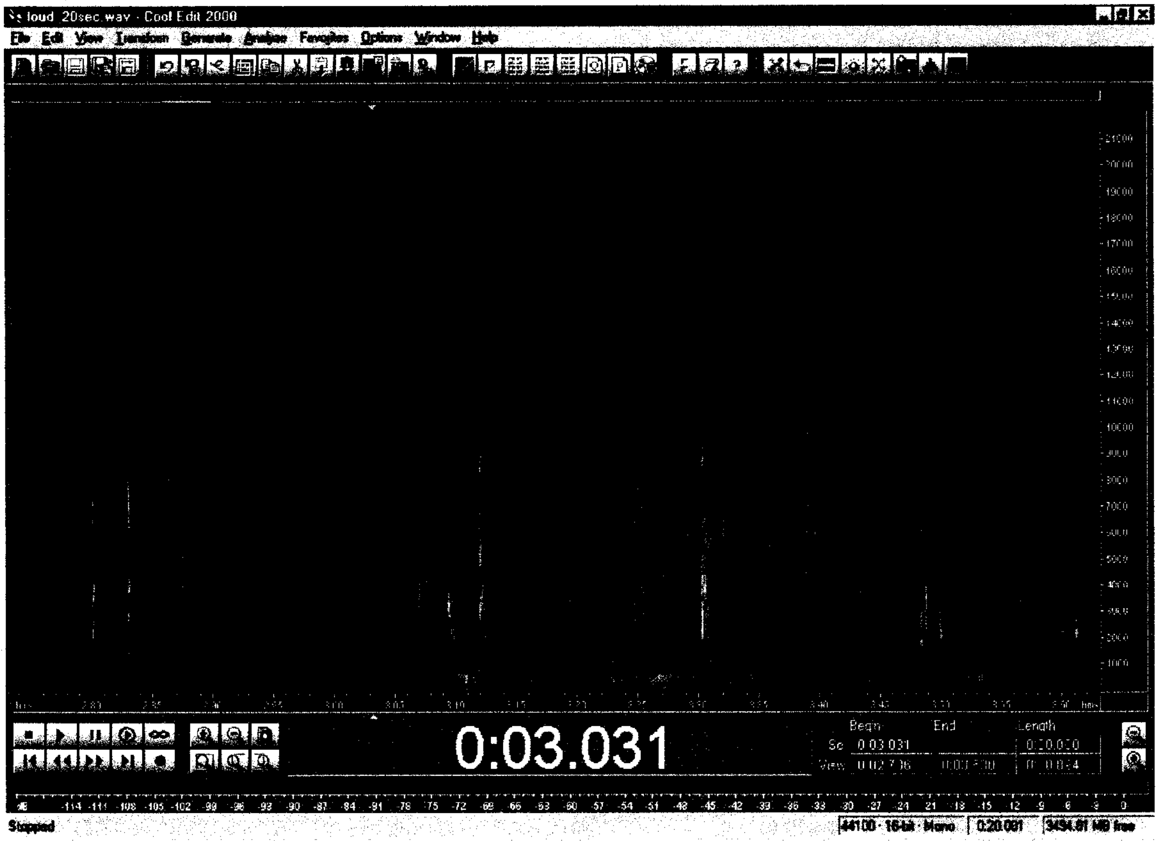
The task of the recognizer is to respond to manatee vocalizations while rejecting non-manatee signals that produce false triggers. There may be a major problem with false triggers on signals such as shrimp clicks that happen, by chance, to occur in groups. Thus, the manatee problem may be summarized as trying to design a recognizer that has a low equal error rate. The equal error rate is achieved when the percentage of misses of manatee vocalizations is equal to the percentage of false triggers due to spurious sounds. Depending on the statistics of the frequency distribution of manatee utterances, it may be acceptable to have a larger error rate on true manatee vocalizations (because, when they occur, they occur in groups, so missing one is not bad) in order to also decrease the false trigger rates. For example, if manatee utterances occur in groups of at least N per minute when they are present, then false triggers may be further reduced by requiring, for example, at least N/2 triggers in a minute before a positive result is reported. This example assumes that the likelihood of N/2 false triggers in a minute is very small. To determine the exact criteria for a positive result, signals in various noise environments must also be available, and knowledge of the frequency distribution of manatee vocalizations when manatees are present must be provided.

If: -original, unmodified manatee recordings in various noise environments are made available

- the manatee detector is tested on original, unmodified recordings or in the water
- statistics on the frequency of manatee vocalizations at times when manatees are present, is provided then, the likelihood that the manatee recognizer described below will function in a satisfactory manner is good.

TECHNICAL APPROACH

The following figure gives a spectrogram of a manatee recording in a background of shrimp clicks.

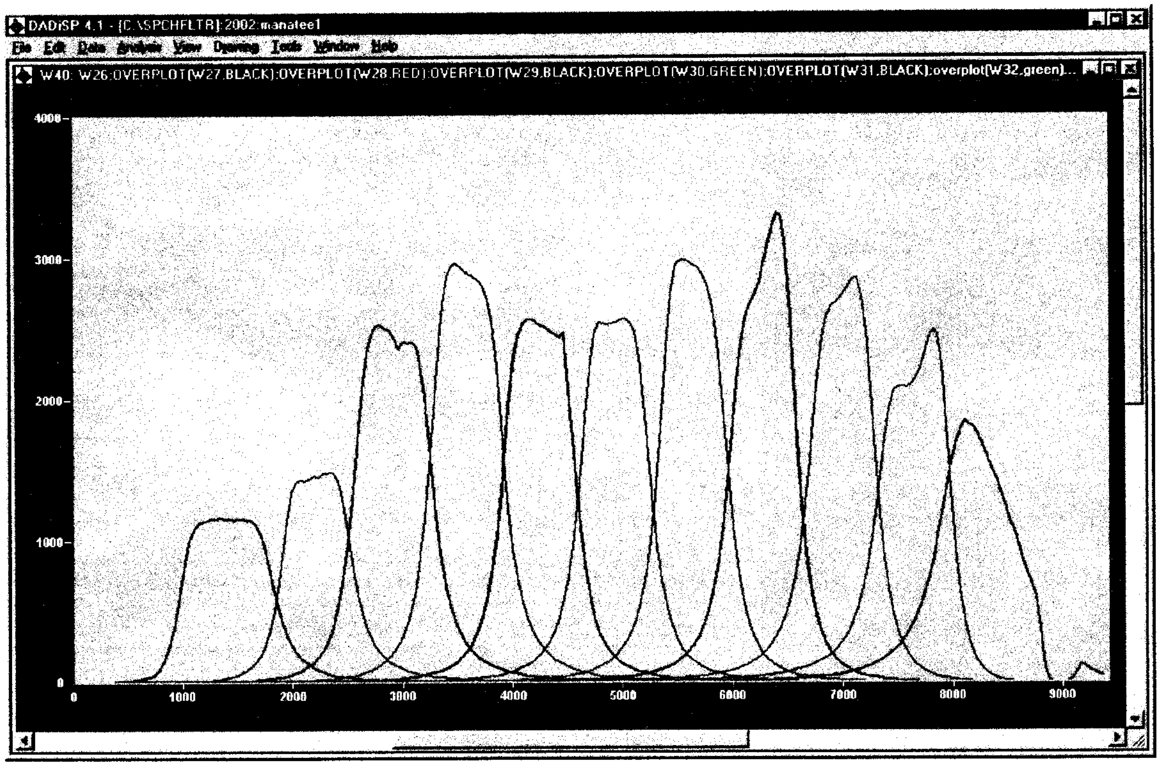


The manatee vocalization occurs about 75% of the way through the recording. It consists of three discrete frequency bands of sound that last for about 150 milliseconds. During this manatee vocalization there are also approximately three shrimp clicks. About 150 milliseconds before the manatee vocalization, there are about 4 or 5 shrimp clicks that are grouped together to produce what might be a false trigger of the Sensory recognizer. Training the neural network in the recognizer to distinguish between the true and false signals is the reason that a large set of examples of each is required.

Because the durations of manatee utterances are typically a few hundred milliseconds, the recognizer will be trained to discard audio signals that are much longer or shorter than this duration.

The neural network is trained on acoustic features extracted in real time by the Sensory LSI chip from the recordings of manatees and the backgrounds. The most important acoustic features of manatee vocalizations are the multiple frequencies, narrow bandwidth emissions, such as those in the above figure. Thus, the ideal acoustic features to extract from the audio signal would be the power outputs of a many channel spectrum analyzer that runs in real time and that covers the frequency range up to ~10 kHz. While this is not the manner in which the Sensory recognizer operates, computer simulations of such a filter bank running in the Sensory

4128 LSI chip have been made. An example of a filter bank that might be implemented in this chip is given below.



The steps in development of the Manatee Proximity Indicator, under the assumption that the data in the overview section is available, would be:

1. Complete the computer simulations of the filter bank.
2. Incorporate this design into the 4128 LSI chip and verify its operation.
3. From the manatee and noise recordings, prepare short segments of manatee vocalizations and all non-manatee sounds that might be heard by the recognizer.
4. Train the neural network on data that is the output of the above filters for both manatee vocalizations and noise.
5. Incorporate a statistical model to handle positive outputs of the recognizer, as described above, in order to minimize the false trigger rate.
6. Report the above results to the funding agency, and test the device against new recordings of original, unmodified manatee vocalizations.
7. Re-design and validate a watertight system for testing the MPI in the actual environment of its anticipated use.

Date Completed: 2/14/03

W. Randolph Warner

cc: Rosa Dalaly
Elsa Haubold
Katherine Frisch