Autonomous Acoustic Trigger for Distributed Underwater Visual Monitoring Systems

Antonella Wilby[1], Ethan Slattery[2], Andrew Hostler[3], and Ryan Kastner[1]

[1]Computer Science and Engineering Dept., University of California, San Diego
[2]Computer Engineering Dept., University of California, Santa Cruz
[3]Electrical Engineering Dept., California Polytechnic State University, San Luis Obispo

ABSTRACT

The ability to obtain reliable, long-term visual data in marine habitats has the potential to transform biological surveys of marine species. However, the underwater environment poses several challenges to visual monitoring: turbidity and light attenuation impede the range of optical sensors, biofouling clouds lenses and underwater housings, and marine species typically range over a large area, far outside of the range of a single camera sensor. Due to these factors, a continuously-recording or time-lapse visual sensor will not be gathering useful data the majority of the time, wasting battery life and filling limited onboard storage with useless images. These limitations make visual monitoring difficult in marine environments, but visual data is invaluable to biologists studying the behaviors and interactions of a species. This paper describes an acoustic-based, autonomous triggering approach to counter the current limitations of underwater visual sensing, and motivates the need for a distributed sensor network for underwater visual monitoring.

CCS Concepts

•Applied computing → Environmental sciences;

Keywords

autonomous monitoring, underwater cameras, acoustic triggering, biological surveys

1. INTRODUCTION

Biological surveys of marine species often rely on collecting visual observations of a species to infer species behavior. In some species with high abundance or known locations, data collection is trivial; in other species, which may have small population numbers or display reclusive behaviors, the task of collecting visual observations becomes time-consuming, and in some cases, impossible. For example, in the case of the vaquita porpoise (*Phocoena sinus*), fewer than 60 individuals remain [3], and little is known about their behavior since they are rarely observed in the wild. In other species, even those commonly observed, particular important behaviors have never been observed or are rarely recorded; for example, the feeding habits and diet of pygmy killer whales are poorly understood due to the lack of documented observations[8]. Additionally, the majority of visual surveys are performed by biologists in the field, which is costly both in terms of person-hours, boat time, and other expenses. These reasons motivate the need for a remote monitoring approach to marine population studies.

The majority of underwater camera systems used in marine species monitoring employ either time-lapse methods[2, 11], record continuously for short periods of time[13], are human-operated[1], or depend on the use of other methods (*e.g.* Fish Aggregating Devices (FADs), bait, sound)[4, 9, 10] to guarantee the presence of a species of interest within close proximity to a camera sensor. None of the existing underwater camera systems surveyed in the literature incorporate a reliable triggering mechanism based on the physical presence of a species under study. This limitation has several consequences: 1) a lot of useless data may be generated with comparatively little useful data; 2) it limits the utility of these systems to places where there is a lot of marine activity and species are guaranteed to be present, *e.g.* a coral reef; or 3) it limits the applicability to a small subset of species that can be attracted to a specific area; for example, pelagic fish are attracted to FADs, but marine mammals will not be attracted to the device and thus cannot be studied by a device that relies on a FAD as an attractant mechanism.

This paper presents an remote monitoring approach to collecting visual data on marine species by leveraging their acoustic vocalizations as an autonomous triggering mechanism for an underwater camera system. Our system uses these acoustic vocalizations to reliably determine the presence of a particular species, recording video only when a species of interest is nearby. This auto-triggered approach increases system deployment times by reducing the amount of unnecessary video-recording time, which is the most power-consuming task, thus saving battery power. It also reduces the amount of unusable data collected by only recording videos of the surrounding area when a species has been detected in close physical proximity to the camera system. We also evaluate the need for a distributed network of autonomous sensor nodes in order to monitor species that are rare, widely-dispersed, and infrequently sighted. Our distributed, autonomously-triggered visual monitoring system facilitates the heretofore unprecedented ability to monitor species that are scarcely-sighted and range over a large area.

WUWNET '16, October 24-26, 2016, Shanghai, China

© 2016 ACM. ISBN 978-1-4503-4637-5/16/10. . . $15.00

DOI: http://dx.doi.org/10.1145/2999504.3001080

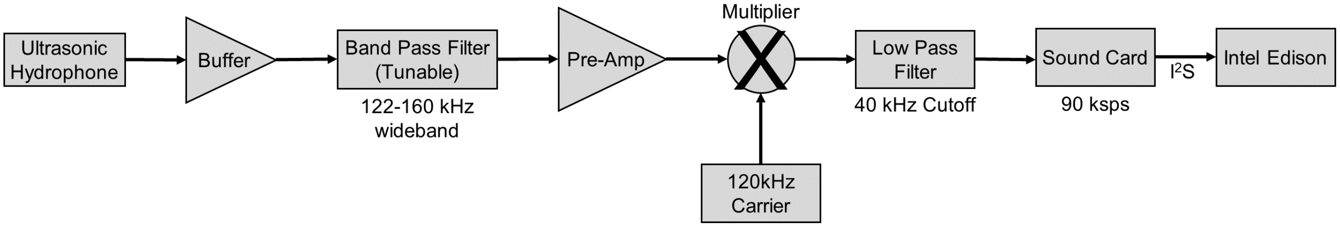

Figure 1: Block diagram of the acoustic triggering system hardware components

2. ACOUSTIC TRIGGERING HARDWARE

Many marine species make vocalizations to communicate with individuals in the underwater environment. Many of these vocalizations, for example the social calls of mysticetes, are in the human-audible frequency range or lower. Other vocalizations, such as the echolocation clicks made by odontocetes for navigation and foraging in their environment, are in the ultrasonic range[7]. Additionally, even species that do not vocalize, such as sharks, are sometimes tagged with data logging tags that emit acoustic pings in order to track individuals. All of these acoustic features can be leveraged to determine an animal's location with respect to the camera system, and used to trigger the cameras if an animal comes within close proximity.

The general method for acoustic triggering is shown in Figure 1. The acoustic signal is first bandpass filtered in hardware to isolate frequencies of interest. If the signals are above 40 kHz, they are modulated to a lower frequency to allow capture with a commercial sound card sampling at the rate of 96 kHz. The raw audio signal is then written to onboard storage and processed to detect vocalizations of interest. Once a signal is detected, one or more cameras are triggered via a GPIO and set to record for a period of time dependent on the distance of the animal to the camera, and whether or not more vocalizations are detected.

Our triggering system uses a Cetacean Research CR3 Hydrophone to detect acoustic signals. This hydrophone was selected for its sensitivity across a wide band of frequencies, with a useable frequency range from 0.1 Hz to 240 kHz. While it does not cover the entire range of vocalizations detected in the marine environment, it covers a large percentage of species.

An Intel Edison computer provides all onboard computation for the system. The Edison was chosen primarily for the amount of computational power it provides in addition to its tiny form factor. Our triggering system hardware is directly integrated with the Edison via a series of custom PCBs that incorporate filtering, amplification, and modulation hardware. These PCBs were designed to keep the same form factor as the Edison, resulting in a miniscule, yet powerful, computational system that fits inside the limited space requirements dictated by the size of the waterproof enclosure housing the triggering hardware.

2.1 Signal Filtering and Amplification

Some environments have a wide variety of species, some vocalizing at low frequencies, others echolocating at ultrasonic frequencies. The ocean's acoustic environment is incredibly noisy at most frequencies, carrying signals from whales as low as 10 Hz, to side scan sonar, ranging as high as 1.6 MHz. These environmental factors must be planned for, and either filtered out or amplified according to their relevance.

Depending on the species of interest, it may be useful to record very wide band frequencies. Many environments have a variety of species, each vocalizing in a different frequency range, and it would be useful for researchers to record the interactions between these species. It is then useful to process a wide range of frequencies to detect multiple species. In case of system or environmental noise, the hardware also incorporates reconfigurable 2nd degree low pass filters. This adjustment allows the system to adapt to acoustic differences in various marine environments.

2.2 Modulation for Ultrasonic Signals

In principle, the acoustic triggering system can be used for any vocalization emitted by a marine species. In practice, as the frequencies of interest shift into the ultrasonic range, the hardware and processing speeds required become nontrivial to implement on low-power embedded systems. In the case of our system, the hydrophone is sensitive to 240 kHz, necessitating sampling at the Nyquist rate of 480 ksps if we wish to utilize the entire frequency range of the hydrophone. This sampling rate is hard to achieve with many embedded systems, which use less power at the expense of computational resources: faster processor speeds and larger storage requirements are needed than many small embedded platforms can provide. To alleviate the sampling rate and storage requirement, modulation hardware was designed to frequency-shift the signals to a lower frequency before digitizing and processing. The power consumption of the entire system including analog filtering and modulation circuitry, sampling and computation is shown in Table 2.

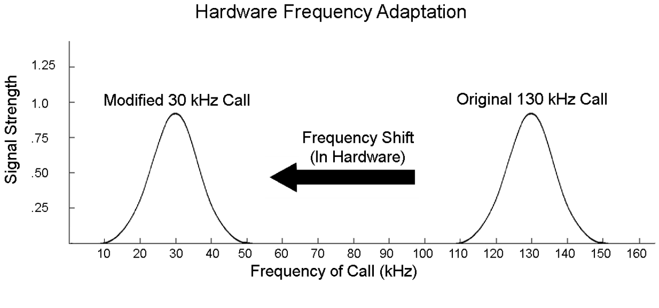

Figure 2: Example ultrasonic signal: frequency-shifted via modulation circuitry from 130 kHz to 30 kHz

The frequency-shifting hardware uses a process that is analogous to AM radio demodulation. The ultrasonic sig-

Section	Power (Sleep)	Power (Active)	Power (Average)
Edison Processor	.24 W	.50 W	.39 W
Analog Hardware	.010 W	.48 W	.28 W
Audio Conversion	.003 W	.005 W	.004 W
Power Supply	.062 W	.08 W	.073 W
Total	**.31 W**	**1.06 W**	**.75 W**

Table 1: System Power Usage across all hardware components

nals from the first stage of the filtering hardware are applied to a carrier wave in a multiplier. The harmonic frequencies output from the multiplier can be computed using:

$$cos(\omega_i) * cos(\omega_c) = \frac{1}{2}cos(\omega_i + \omega_c) + \frac{1}{2}cos(\omega_i - \omega_c) \qquad (1)$$

where ω_i is the unmodulated input signal and ω_c is the 120 kHz carrier wave.

This signal is then highpassed to remove power supplier noise and the lowest harmonic frequencies from the multiplier, then goes through an envelope detector with 3rd degree low pass filtering. The result of this signal transformation is approximately the lower frequency of the two end terms in equation 1. The low pass filter passes the harmonic frequencies from the multiplier that are $\leq$ 40 kHz, and the resultant frequency-modulated signal is then amplified to either microphone or line level.

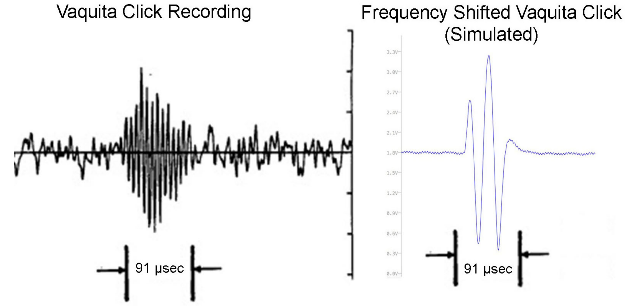

Figure 3: Example of a modulated click: vaquita porpoise click with peak energy at 139 kHz[12], frequency-shifted to 40 kHz

3. ANALOG TO DIGITAL CONVERSION

Once the analog signal has been filtered and modulated, the analog signal must be digitized so that it can be processed and stored on the Edison. Since the modulated analog signal has a maximum frequency of 40 kHz, the analog signal can be digitized with a much slower ADC than would be required to digitize the raw signal. The modulation hardware reduces the minimum required sampling rate from 480 ksps to 80 ksps.

Oversampling will provide headroom for processing tasks. At 20% oversampling the desired sample rate will be 96kHz, which is a standard audio capture frequency which achieves the oversampling goals and will make integration with existing audio processing code much easier. The bit depth of each sample affects the Signal to noise ratio of an ideal ADC according to Equation 2 [6] and several different bit-depths can be seen with their corresponding signal to noise ratios in Table 3.1. A 12-bit sample depth achieves a theoretical signal to noise ratio of 72.25 and roughly aligns with the signal to noise ratio of the analog signal so was selected as the minimum bit-depth.

$$\text{SNR} = 6.02N + 1.76\,\text{dB} \qquad (2)$$

3.1 Discrete Analog to Digital Converters

There are many analog to digital converters available offering a range of sampling rates, bit depths, and communication protocols. Many claim sample rates in the MHz range, but generally as speed increases the bit depth of each sample quickly falls. An external ADC also required the use of a real-time coprocessor to control and read data from the ADC since code running inside the Edison can not meet real-time requirements.

Bit-Depth	SNR	Comments
8	48.16 dB	One Byte
12	72.24 dB	Project Minimum
16	96.32 dB	Standard CD-Audio
24	144.49 dB	DVD or BluRay Audio
32	192.64 dB	Professional Audio Mastering

Table 2: Common ADC Bit-Depths and SNR

Several ADC integrated circuits were selected that exceed the project's minimum requirements, but none proved fruitful in implementation. The design of a circuit to get a noise free signal into the converter, read by the real-time processor, and then transmitted to the Edison with sufficient data bandwidth is a non-trivial task. Details of these trials are out of the scope of this paper, and although it is an interesting problem, there are better solutions available that meet the requirements of this project.

3.2 Integrated Audio Codec Circuits

An 80 kHz sample rate falls comfortably within the range of high-quality consumer audio capture cards which offer sampling rates of 96 kHz. The use of a consumer audio capture device significantly reduces the complexity of the design since they integrate noise reduction, high bit-depth sampling, timing, and transmission protocols all into a single low-cost integrated circuit.

The selection of an audio capture device hinged on the sample rate, bit-depth, and also driver availability for the Linux kernel used by the Intel Edison. The WM8731 audio codec provides 24-bit samples at a rate up to 96 kHz. It can be controlled directly from the Edison via I2C and transmit captured samples to the Edison via I2S while consuming 4.95mW. Linux drivers for the WM8731 exist, and it has been successfully integrated into other embedded Linux

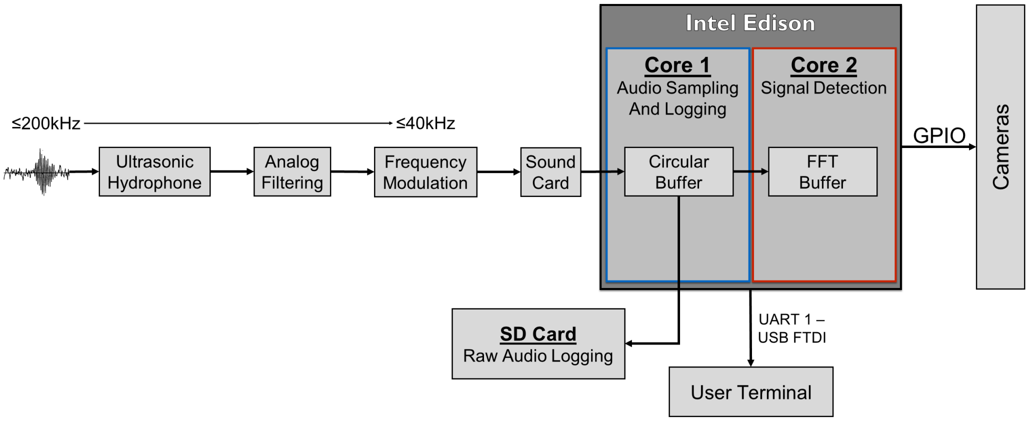

Figure 4: Block diagram of the entire triggering, sampling, and detection system, from hardware to software

systems before. The manufacturer of the WM8731 sells a prototyping board that is a near fit for the form factor of the Edison stack as well, which makes prototyping and deployment without a custom board possible. We can then add this board to our setup easily, while simplifying the design process.

4. SYSTEM SOFTWARE DESIGN

Since an audio codec is used to capture the signal, the stream of data can be processed as normal audio on the embedded computer. Treating the incoming signal as audio enables the system to use the Advanced Linux Sound Architecture (ALSA) libraries, which is common on Linux systems. This section will describe the capture and processing details of the captured signal, referred to as the audio from here forward.

4.1 Audio capture

The WM8731 codec is a programmable audio capture and output device and therefore requires configuration by the Edison before it can begin the capture process. The Edison issues pre-defined I2C commands to request a 96kHz sampling rate and to start the recording process. This configuration step is abstracted away by the kernel drivers but is also possible to do manually in code.

Once sampling begins, the ALSA framework makes available the incoming samples. To make interfacing with ALSA more streamlined the system utilizes a popular open source library PortAudio. PortAudio allows the easy configuration of incoming audio streams, their properties, as well as callback functions to handle incoming data. PortAudio working with ALSA handles the interrupts created by the incoming data. The callback function implemented for this system places new samples in a large circular buffer so that the audio data continues uninterrupted while the storage and signal processing sections of the code act as consumers.

4.2 Audio Storage

Storing raw audio is not required for the triggering of the cameras in this system, but it is none the less an important part of the system as a whole. Many species in the ocean lack quality and openly available sample data of their calls and echolocation clicks. By recording raw audio to a storage medium, the system can help improve the collection of bioacoustic data available for detection and analysis.

Large capacity microSD cards are available on the consumer market in 64GB, 128GB, and even 256GB sizes. To enable the audio capture to last as long as possible the memory available to store recordings needs to be as large as possible. The large size of the SD necessitates a file system that can handle large volumes. The standard exFat file system has a proprietary license, and so it does not suit the system. Since the system is running a Linux kernel though it does support ext4 which can handle volumes and files up to 16TB, which is far beyond this system's requirements.

microSD Size	24 bit	16 bit
64	61.7 Hours	92.6 Hours
128	123.5 Hours	185.2 Hours
256	246.9 Hours	370.4 Hours

Table 3: Hours of audio storage at 96 kHz sample rate over different sizes of microSD cards

A non-trivial problem when writing to an SD card is the latency of the write operation. Most SD cards advertise up to 60 MB/sec, but this is only for bulk data transfers. Writing samples to the card as they arrive could result in lost data as the write operations quickly fall behind. The SD specifications state that bulk writes happen in blocks of 512 bytes, therefore writes that are a multiple of 512 will be most efficient. Therefore a small circular buffer of 512 byte blocks is used to ensure that every write operation executes as quickly as possible.

4.3 Frequency Detection and Triggering

In parallel to the sampling and storage of the audio data, the data is also continuously processed to detect signals of interest, in order to trigger the cameras when a target species is detected. Currently, the system detects the narrowband, high-frequency clicks produced by porpoise species, using the amplitude of the peak energy frequency to trigger the camera system. More sophisticated signal detection algorithms could be produced using training datasets for specific species.

Audio samples are transformed from the time domain to the frequency domain using a Fast Fourier Transform (FFT).

The audio samples are a one-dimensional array of purely real data so the output of this transformation is an array of complex values that represents the magnitude and phase of the signal. Extracting the magnitude of a certain signal is a simple matter of, making sure to take into account the shift induced by the analog circuitry, dividing the capture bandwidth by the number of bins in the transformations. Each bin will cover a certain bandwidth of the captured spectrum and monitoring the magnitude of the correct bin will alert the computer to the presence of a signal in that frequency range.

The library Fastest Fourier Transform in the West (FFTW) [5] was used to perform FFT transformations because it is open source, adapts and optimizes to the hardware it is run on, and accepts a wide variety of data input formats [5]. Each audio sample coming from the sound card is a signed integers and so is first transformed into a 32-bit floating point value between -1.0 and 1.0. Not only is this standard for audio recording but the FFT runs much faster on floating point data. When using single precision data FFTW performs transforms fastest on windows of 256 samples [5] so a sliding window of that size is used, with overlaps of 128 samples.

5. DISTRIBUTED SYSTEM

The autonomous triggering system described is currently implemented on a single prototype, deployed in the Sea of Cortez in Mexico to monitor the vaquita porpoise. It is attached to a mooring on the ocean floor and floats below a buoy, strategically deployed in an area where vaquita are frequently detected by existing acoustic recorders deployed in the area. Although this system can gather data for a week or more, our single prototype is limited to the detection range of the high-frequency acoustic clicks, around 250m or less.

The use case of the vaquita species especially motivates the need for a distributed system. Since so few remain, and since so little is known about the species, a distributed network of nodes provides the best probability of getting visual observations of the species, which in turn is critical to the conservation effort. In other species, particularly those that are rare or range over large areas, a single monitoring station is ineffective, producing at best infrequent observations of individuals. This autonomous acoustic trigger can be integrated into a distributed network of visual monitoring nodes, in order to provide more frequent observations of multiple individuals in a population under study. Monitoring efforts of many marine species will also benefit from a distributed approach, resulting in more data of more individuals interacting over time, allowing biologists to gain clearer insight into the behaviors of little-understood species.

6. CONCLUSIONS

Our system improves the state-of-the-art of underwater monitoring systems by introducing autonomous triggering based on acoustic presence indicators of marine species. This approach provides an improvement over the current state-of-the-art by incorporating real-time response to acoustic stimuli and a low-power system for ultrasonic signal processing, coupled with visual monitoring capabilities. We introduce a hardware method for signal filtering and modulation, and the hardware and software necessary to digitize and detect cetacean vocalizations in an audio signal.

This system has wide applicability in many fields of ocean science and marine biology, including behavioral studies of rare and endangered species, marine population census, and long-term habitat monitoring. The system has the potential to produce data on species and specific behaviors that have never been observed in the wild before, and reduce the amount of useless data generated while simultaneously saving power and increasing deployment times via the autonomously-triggered approach.

7. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under REU Award #1560162, the UC San Diego Frontiers of Innovation Scholars Program, and the National Geographic Society under grant #C292-15.

8. REFERENCES

[1] M. Benfield. Zoovis: a high resolution digital camera system for quantifying zooplankton abundance and environmental data. *ASLO Aquatic Sciences Meeting*, Feb. 2001.

[2] F. Cazenave. Seestar: A low-cost,modular and open-source camera system for subsea observations. *IEEE OCEANS '14*, pages 1–7, Sept. 2014.

[3] CIRVA. *Seventh Meeting of the Comité Internacional para la Recuperación de la Vaquita*, Ensenada, BC, Mexico, May 2016.

[4] R. Driscoll. Species identification in seamount fish aggregations using moored underwater video. *ICES J. Mar. Sci.*, 69(4):648–659, Feb. 2012.

[5] M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".

[6] W. Kester, editor. *Data Conversion Handbook.* Newnes, Burlington, MA, 2005.

[7] D. R. Ketten. The marine mammal ear: Specializations for aquatic audition and echolocation. *The Evolutionary Biology of Hearing*, New York: Springer:717–750, 1992.

[8] D. J. McSweeney, R. W. Baird, S. D. Mahaffy, D. L. Webster, and G. S. Schorr. Site fidelity and association patterns of a rare species: Pygmy killer whales (feresa attenuata) in the main hawaiian islands. *Marine Mammal Science*, 25(3):557–572, 2009.

[9] H. M. Murphy and G. P. Jenkins. Observational methods used in marine spatial monitoring of fishes and associated habitats: a review. *Marine and Freshwater Research*, 61(2):236–252, Feb. 2010.

[10] A. A. Myrberg. Shark attraction using a video-acoustic system. *Marine Biol.*, 2(3):264–276, Mar. 1969.

[11] A. Sherman. Deep-sea benthic boundary layer communities and food supply: A long-term monitoring strategy. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56(19-20):1754–1762, Sept. 2009.

[12] G. Silber. Acoustic signals of the vaquita (phocoena sinus). *Aquatic Mammals*, 17(3):130–133, June 1991.

[13] A. Stoner. Flatfish-habitat associations in alaska nursery grounds: Use of continuous video records for multi-scale spatial analysis. *J. Sea Research*, 57(2-3):137–150, Feb. 2007.