Design of a Low-Cost and Extensible Acoustically-Triggered Camera System for Marine Population Monitoring

Antonella Wilby, Ryan Kastner
Dept. of Computer Science and Engineering
University of California, San Diego
La Jolla, California 92093-0404
{awilby, kastner} @eng.ucsd.edu

Andrew Hostler
Dept. of Electrical Engineering
California Polytechnic State University, San Luis Obispo
San Luis Obispo, California 93407-0355
ahostler@calpoly.edu

Ethan Slattery
Dept. of Computer Engineering
University of California, Santa Cruz
Santa Cruz, California 95064
eslatter@ucsc.edu

***Abstract*—As the health of the ocean continues to decline, more and more marine populations are at risk of extinction. A significant challenge facing conservation biologists is the ability to effectively monitor at-risk populations due to the challenges of the underwater environment. Obtaining visual data on a marine species typically requires significant time spent by humans observing in the field, which is both costly and time-consuming, and often yields a small amount of data. We present a low-cost, acoustically-triggered camera system to enable remote monitoring and identification of marine populations.**

I. Introduction

Marine population studies require significant time spent observing in the field. This field time is costly, both monetarily and in terms of person-hours, and often yields relatively little data, especially in cases where a particular population of interest is very small or reclusive. For this reason, much of marine monitoring focuses on acoustic sensing as the primary mode for information gathering, since acoustic sensors can be deployed long-term without need for a human in the field. However, behavioral analysis requires visual observation of multiple individuals of a species interacting over time, and acoustic monitoring alone cannot provide enough information for a complete analysis. Understanding the behaviors of at-risk species is often key to implementing effective conservation strategies[1].

Optical sensors are often ignored as a mode for data collection in the ocean environment because the characteristics of the underwater environment (*e.g.* turbidity, light attenuation, etc.) limit their utility. However, the need to obtain visual data in order to analyze the behavior, identify individuals, or monitor the health of a species suggests that optical sensing could have wide-ranging applicability in marine observation systems.

We present a system that couples acoustic and optical sensors, leveraging the innate acoustic signatures of certain marine species as the triggering mechanism for an autonomous camera system. Our system gives biologists the ability to acquire visual data on these species, without the overhead of significant field time. Additionally, the acoustic trigger provides a reliable indicator of the physical presence of a particular species, removing the element of chance of a timelapse-based system, while consuming less power and creating less useless data than a continuously-recording system.

A. Related Work

Most existing underwater camera systems for species or habitat monitoring incorporate no automatic triggering mechanism that is based on presence indicators of a species of interest. Instead, these existing systems rely on time-lapse techniques (which are not guaranteed to capture any subject of interest), human control, or methods for attracting species to the camera. Many of these systems are also deployed in areas that are abundant with marine life, such as coral reefs, guaranteeing that the system will capture images of some species, if not a particular species of interest.

The rest of this section presents a survey of existing underwater camera systems applied to marine species monitoring that are either time-lapse-based systems, attraction-based systems, or human-operated acoustic-video systems.

1) Timelapse-Based Video Monitoring Systems: Many low-cost underwater visual remote monitoring systems use a simple timelapse method as the triggering mechanism for the camera sensors. SeeStar, the low-cost, modular camera system described in [2], can be integrated onto ROVs or AUVs or tethered on a mooring, and is programmed to take photos at regular intervals. The time-lapse camera system of [3] takes hourly photos of 20 m^2 of the seafloor for deep-sea benthic community studies. The major downside of timelapse techniques, especially in studies where the population of interest is sparsely distributed over an area, is that a timelapse camera is not guaranteed to capture any data ("zero counts"), even over a long deployment period.

2) Attraction-Based Video Monitoring Systems: Other underwater visual observation systems use methods of attracting marine species to a base station where a camera system is deployed. The video systems described in [4] and [5] use Fish

Aggregating Devices (FADs) as an attractor to guarantee the presence of pelagic fish near the camera systems (both towed and moored, respectively). The underwater video camera used in [6] leverages the fact that several species of sharks have been shown to be attracted to low-frequency pulsed sounds, and used irregularly pulsed signals below 1000Hz to attract sharks to their moored camera system. Other systems, such as those described in [7] and [8], use bait such as pulverized fish to attract species and reduce zero counts.

3) Human-Operated Monitoring Systems: The most common underwater visual observation method is human-based, whether a diver in the water with a camera, such as in [9], or human-in-the-loop, where a camera system is remotely operated by a human, such as in [10].

B. Background

In this paper, we will present the use case of the endangered vaquita porpoise (*Phocoena sinus*) to motivate the need for our acoustically-triggered approach to underwater camera systems. In the case of the vaquita, their population is incredibly small, around 60 individuals, so the probability of a timelapse system being effective is minute. It is not known whether vaquita are attracted to bait or sound, so attraction-based systems may be ineffective. The vaquita are very reclusive, fleeing from the sound of approaching boat motors, making a human-in-the-loop approach ineffective[11]. Other marine species besides vaquita fit these characteristics as well, and those populations could also be served by a novel monitoring approach.

Vaquita and the other odontocetes use echolocation for navigation and finding food, and are the only marine mammals that are known to echolocate. The echolocation clicks made by odontocetes are divided into two types: Type 1, echolocation clicks with peak spectra above 100 kHz, and Type 2, with peak spectra below 80 kHz[12]. The vaquita's echolocation click ranges from 122 kHz to 146 kHz, with peak energy at 139 kHz [13]. Other marine animals, such as the mysticetes, pinnipeds, and some species of fish, make audible vocalizations for communication. Additionally, many species are tagged by biologists with acoustic tags, which collect data on the animal and also transmit acoustic pings. Echolocation clicks, audible vocalizations, and acoustic tag pings all provide a reliable and distinguishable indicator of a particular animal's presence.

C. Contributions of this Work

Our work introduces a low-cost optical sensing system that leverages these acoustic signatures as a triggering mechanism to reliably capture visual data of marine animals. This system provides a novel way to couple optical and acoustic data in a marine environment, in order to remotely monitor at-risk marine populations, such as vaquita. The major contributions of this work are:

- A novel camera triggering method that leverages acoustic signatures of marine species as the trigger signal.
- A spherical camera system with high enough resolution to resolve characteristic markings in order to differentiate individuals of a species.
- Design of a low-cost, extensible platform for acoustically-triggered data gathering, which can easily integrate many other sensors to provide comprehensive ocean sensing.
- A powerful computing system that can be reprogrammed to capture data on multiple species or perform onboard data processing.

Thus far, our system design has focused on detecting only high-frequency (Type 1) echolocation clicks. However, we are currently generalizing the system to detect Type 2 clicks as well as audible vocalizations and pings, which is discussed further in section VII. For the remainder of the paper, we will refer to "vocalizations," "acoustic signals," or "acoustic signatures" as general terms encompassing both types of echolocation in addition to vocalizations, in accordance with our eventual goal of making the system a species-invariant marine monitoring platform.

II. System Overview

Our system leverages the acoustic signatures of marine animals as a reliable triggering mechanism for obtaining visual data on species of interest. The system integrates a hydrophone as a trigger, an onboard computer which performs sampling, acoustic data logging, and signal processing, and six cameras which provide a 360°by 360°field of view and record 1080p video when triggered.

Fig. 1. Working Prototyping of Camera System

The system detects acoustic signals via an ultrasonic hydrophone. The analog signal from the hydrophone is run through analog filtering circuitry, then digitized and processed on an onboard microcontroller. If a signal from a species of interest is detected within some distance threshold, the cameras are triggered to record for a period of time that scales with the magnitude of the distance at which the species was detected.

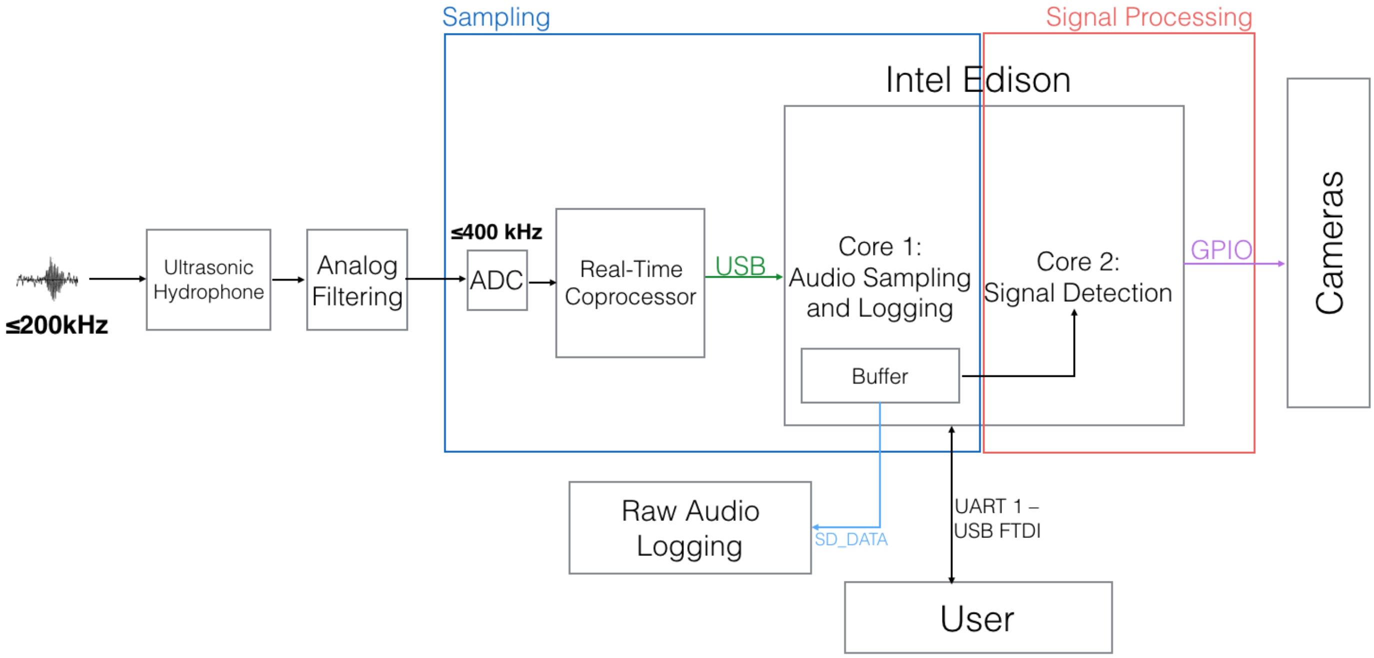

Fig. 2. Block Diagram of Major Systems

A. System Cost

The system is intended to be as low-cost as possible, in order to remain an accessible monitoring tool for ecologists, conservationists, and biologists. The current cost of manufacturing the system is under \$5500, with approximately half the cost coming from the hydrophone and cameras. This cost could be reduced if the acoustic signals of interest are in the human-audible range, by replacing the ultrasonic hydrophone with a lower-cost hydrophone with a smaller frequency range. Similarly, the cameras could be replaced with higher-quality (and thus higher-cost) cameras if the application requires higher-resolution imagery.

Fig. 3. Intel Edison Platform, with breakout boards for console, GPIO, microSD, and ADC (custom boards not pictured)

B. Onboard Computation

The onboard computation platform is the Intel Edison, chosen primarily for its combination of high performance and tiny form factor. The Edison uses Intel's x86 architecture and runs embedded Linux. The CPU is a dual-core Intel Atom which runs at 500MHz. In our current implementation, one thread samples, buffers, and logs the acoustic data to a microSD card, and a second thread reads the buffer, processes the signal, and triggers the cameras if a detection is made.

The Edison (Figure 3) is designed to accommodate a variety of breakout boards of the same or similar form factor to the Edison itself. We have implemented a set of custom breakout boards, which perform analog filtering and ADC sampling. The ADC board breaks out the Edison's SPI interface, and the ADC sends the data to the CPU over this interface. The raw audio is logged to a microSD using the Edison's microSD breakout board. The input to the ADC can either be the raw acoustic data from the hydrophone, in which case the ADC samples at 300 kHz, or the output of the analog filtering circuitry, discussed further in section III.

C. Power Consumption

Table I shows the total power consumption of the system. The cameras and computer consume the most power, and the power consumed by other components is negligible. The system shuts down for 10 hours at night, to reduce power consumption, and because the system relies on natural light instead of artificial lights. We estimate that only at most 5% of the time the system will actually be recording video, which is the most power-consuming task, while the remaining time the system will be logging and processing audio. However, future field experiments, as well as the abundance of the particular species being monitored, will inform that parameter.

TABLE I
TOTAL SYSTEM POWER CONSUMPTION

	State		
	Powered Down	Audio Processing	Recording Video
% Time in State (Est.)	42%	53%	5%
Current Draw (mA)	0.5	200	2362
Power (mW)	1.7	660.4	9660
Total Power Consumption (mW)	**847 mW**		

The system is powered by a 12V battery pack that can supply a maximum of 151 Watt-hours. With a total power consumption of 847mW, which assumes recording video only 5% of the time, we get approximately 180 hours of runtime. Our initial design requirements specified that we needed one week of runtime, because we anticipated that a deployment of longer than one week would result in poor images due to the onset of biofouling[14]. However, more experimentation is needed to determine if the system can be deployed for longer than a week without drastically reducing image quality, and if so, a larger battery could be incorporated so that system maintenance time is reduced.

D. Mechanical Design

The mechanical design is modular (Figure 4). Each camera is contained in a separate housing, and the batteries, computers, and power electronics are contained in another internal housing. Each camera housing is connected to the internal housing with an underwater cable, which provides both power from the system battery and a trigger signal from the computer. Each camera has an individual voltage regulator also contained in the camera housing.

The internal housing contains the 12V battery pack, the Intel Edison computer, and the power electronics (including power distribution board and the voltage regulator for the Edison). The main switch is outside the chassis of the system.

The modular design was chosen over a single-housing design to minimize component loss in the case of housing failure.

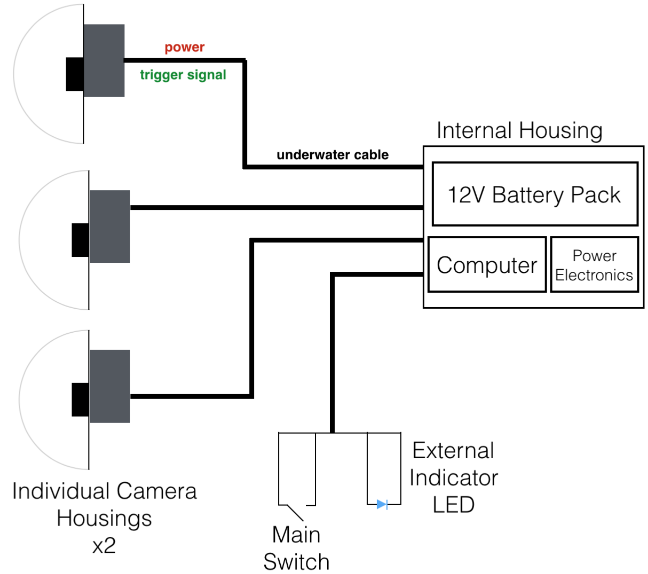

Fig. 4. Modular mechanical design of camera system, with each component housed in a separate enclosure

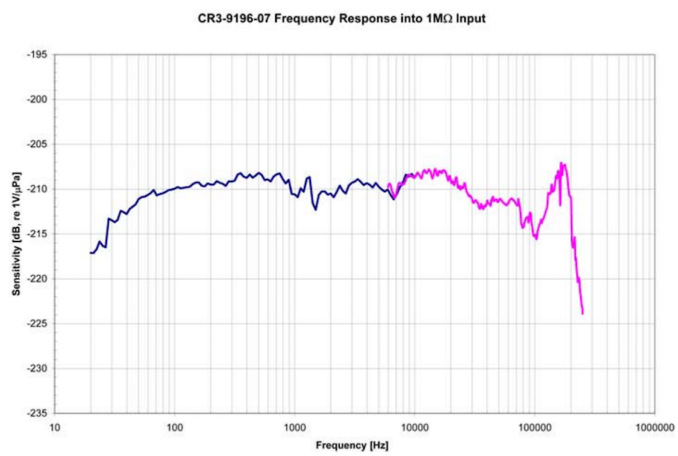

Fig. 5. Frequency response of CR3 Hydrophone from Cetacean Research Technologies, taken from[16]

III. ACOUSTIC TRIGGERING SYSTEM

Echolocation clicks and vocalizations of marine mammals are reliable indicators of the presence of a particular species in a survey area. These acoustic signatures are often used to do population counts and other surveys[15]. Using echolocation as a trigger presents a particular engineering challenge because the clicks are in the ultrasonic range, requiring a faster Analog to Digital (ADC) converter and faster processing and write speeds to sample and record the signal.

The omnidirectional hydrophone used in our prototype is sensitive from 20 Hz up to approximately 200 kHz (Figure 5). Due to this wide frequency response, it can detect audible vocalizations, high-frequency echolocation clicks, and pings from many types of acoustic tags.

The triggering system can be configured in two different ways (Figure 6). In the first configuration, designed to capture audio at the Nyquist frequency of the raw signal, the hydrophone signal is directly sampled by the ADC at speeds up to 400kHz sampling rate. In the second configuration, the signal is first fed into an analog filtering circuit which bandpasses the signal, amplifies it, multiplies it with a carrier frequency, low-passes the signal, and is finally sampled by the ADC (Figure 6).

A. Analog Filtering and Sampling

Figure 7 shows the design of the analog filtering circuitry. The primary purpose of this circuitry is twofold: 1) to isolate the frequency band of the acoustic signals of interest using the bandpass filter, and 2) to modulate the signal to a lower frequency in order to perform sampling, recording, and signal processing at a lower frequency.

The signal is first buffered in order to match the impedance of the hydrophone with the impedance of the circuit, and to

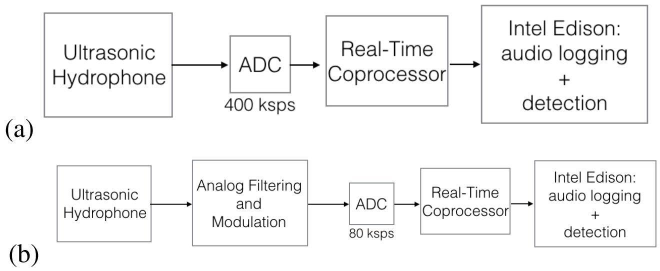

Fig. 6. Two configurations of acoustic recording and triggering system. (a) is used to record and process the entire 20 Hz to 200 kHz band of acoustic data at the Nyquist rate of the signal, removing the step of modulation. (b) is used to process only the desired frequency band and modulate the signal to a lower frequency for reduced power consumption.

convert the passive hydrophone signal into an active one to maintain signal integrity throughout the system. Then, the signal is bandpassed to keep only the desired frequency band. The bandpass filter is tunable in order to accommodate acoustic signals in different frequency ranges, allowing us to detect both types of echolocation clicks, audible vocalizations, and pings from tags. For example, the frequency range show in Figure 7 is the correct range for monitoring the vaquita's ultrasonic echolocation clicks. Since the hydrophone is sensitive from 20 Hz to 200 kHz, the bandpass can be tuned to pass a different frequency band depending on the frequency of the specified vocalization.

The signal is then amplified, then modulated using a 120kHz carrier signal. Finally, the signal is run through a low pass filter which passes the harmonic frequencies from the modulated signal that are below 40 kHz. This allows the triggering ADC to sample at the Nyquist frequency of a 40 kHz signal of 80 ksps (kilosamples per second), reducing both power consumption and the computational time necessary to sample and process the signal. Sampling at this lower rate also reduces the memory storage necessary for the ultrasonic audio, allowing for potentially lower-cost memory as this signal dataset would take approximately $\frac{1}{4}$ the space on a microSD card.

B. Signal Processing and Detection

The filtering circuitry makes the task of signal processing much simpler. Without the filtering system, for signals from 122 to 160 kHz (the maximum frequency the system is currently designed to sample), the ADC must sample at the Nyquist rate of 320 kHz, and the signal detection algorithm must process the incoming data at that frequency. Instead, we modulate the signal to 40 kHz and only sample and process the signal at 80 kHZ, resulting in reduced computation time and power consumption.

Sampling and detection are implemented in two separate threads, each running on a core of the Intel Edison CPU. The sampling algorithm buffers the incoming samples using a double buffer—these buffers are then written both to the microSD card which stores the raw audio, and sent to the signal detection algorithm.

The current detection algorithm uses a cross-correlation matched filter to match the known vaquita signal [13] in the noisy incoming signal.

We are currently investigating detection algorithms for other species which will use a machine learning approach to train the detector, specifically artificial neural networks trained on real acoustic data. The reason this approach was not taken for the vaquita use case was we did not have access to a large enough dataset of real vaquita vocalizations to train a neural network.

IV. Optical Characteristics

The primary sensor payload is a spherical camera system consisting of six cameras with fisheye lenses, which provides a 360°by 360°field of view. Each camera is inside a low-cost, acrylic dome port, which preserves the camera's complete field of view by eliminating the flat-port housing interface with the water. The spherical camera will detect any specimen within visible range, regardless of orientation with respect to the camera. Combined with the omnidirectional hydrophone, this spherical camera approach eliminates the need for an actuated system, reducing both power consumption, cost, and complexity.

The camera provides enough resolution to resolve characteristic markings on species of interest, in order to differentiate individual members of a species. This allows biologists to monitor the health of individuals over time, if individuals are detected by the camera system multiple times.

The theoretical minimum resolvable feature size for 1080p video recording was computed using Eq. 1.

$$P = \frac{h_p}{24 * d * tan(FOV_h)} * l_t * \frac{v_p}{24 * d * tan(FOV_v)} * w_t \quad (1)$$

where:

$P =$ number of pixels on the target feature
$h_p =$ horizontal pixels on the sensor
$v_p =$ vertical pixels on the sensor
$FOV_h =$ horizontal field of view in radians
$FOV_v =$ vertical field of view in radians
$d =$ distance to target in feet
$l_t =$ length of target feature in inches
$w_t =$ width of target features in inches

An identifiable feature that is 3 inches by 1 inch will be resolved by approximately 24 pixels at a distance of 15 feet when recording 1080p video. Features smaller than 0.5 inches square will only have two pixels on target. In our experiments, at least 15 pixels were required on target to be able to differentiate printed letters, which corresponds to a minimum feature size of approximately 1.5 inches by 1.25 inches.

For photos, the resolution improves slightly. Each camera has a 5 Megapixel sensor, which gives 59 pixels on target for our reference target feature of 3 inches by 1 inch at a distance of 15 ft from sensor. Using the metric of minimum 15 pixels

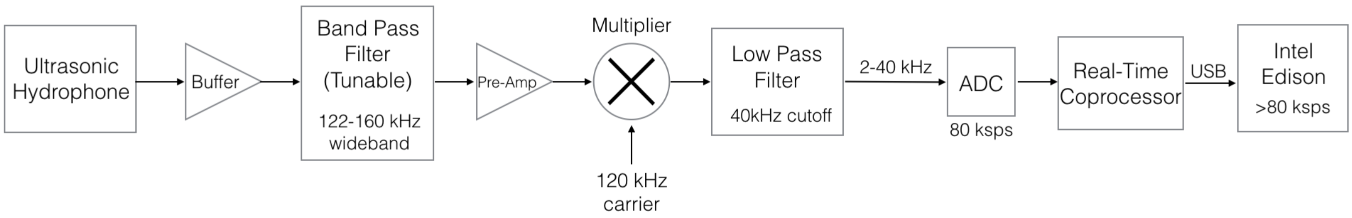

Fig. 7. Analog Filtering Circuitry

per target, if the system is taking still photos the minimum resolvable feature is 1 inch by 0.75 inches.

However, these calculations do not take into account that the placement of identifiable markings contains as much identifiable information as the shape and characteristics of the markings themselves. Thus, the low resolution may not impede the differentiation of individuals with different spacial distributions of identifying markings in many cases.

The optics of the system were characterized using Modulation Transfer Function (MTF) testing, specifically using an open-source software called MTFMapper [17]. The Modulation Transfer Function is a measure of the sharpness of an image, measured in cycles per pixel, which is the spatial frequency at which information can be resolved by a particular imaging system. The ideal MTF corresponds to the Nyquist frequency at 0.5 cycles/pixel. In practice, 0.25 cycles/pixel to 0.33 cycles per pixel is considered an acceptable image sharpness[18].

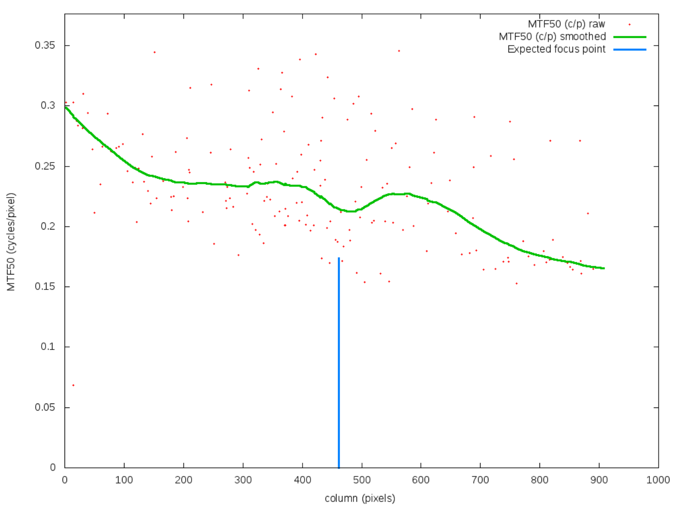

Fig. 8. MTF values across the columns of the image plane, with complete lens and dome setup

Figure 8 shows the MTF values as they vary across the columns of the image plane. In this graph, it appears that our image is sharper at the left side of the sensor, and decreases in sharpness at the opposite side. Figure 9 shows the MTF values from testing just the camera lens, without a dome, and the image is much sharper across the entire image plane. These differences can likely be accounted for by manufacturing variations in the low-cost acrylic domes. However, the results obtained are still above the acceptable threshold of 0.25 cycles/pixel.

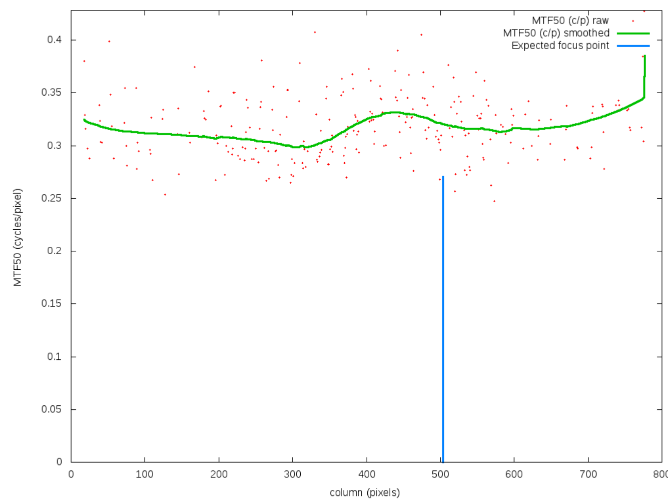

Fig. 9. MTF values across the columns of the image plane, with just the lens and without dome

Figure 10 shows a two-dimensional visual representation of the lens and dome acuity as it varies across the sensor. The plot in (a) shows the MTF values across meridional lines, which are lines in planes that include the optical axis, also known as axial lines. Green denotes an area of greatest sharpness, around 0.35 cycles/pixel, whereas blue and white are the areas of least sharpness, from 0.2 down to 0.15 cycles/pixel. Plot (b) shows the same for the sagittal lines, or radial lines.

Figure 11 shows a three-dimensional representation of the same MTF values across the focal plane, for both meridional and sagittal lines. As in the two-dimensional representation, green denotes the sharpest areas while white denotes the areas of least sharpness.

These results demonstrate that the choice of low-cost acrylic domes is degrading the quality of the images obtained, although in practice we have not noticed a decreased ability to resolve the features of interest.

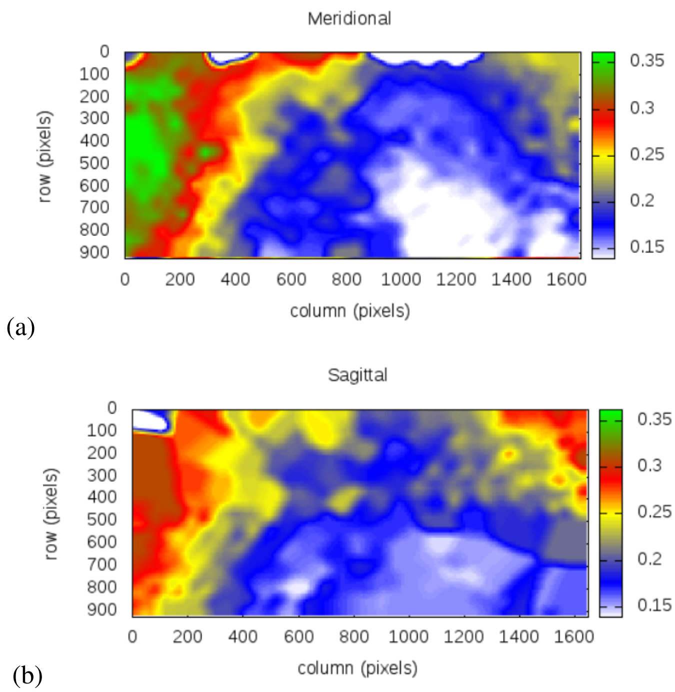

Fig. 10. 2D representation of MTF values for camera dome and lens system

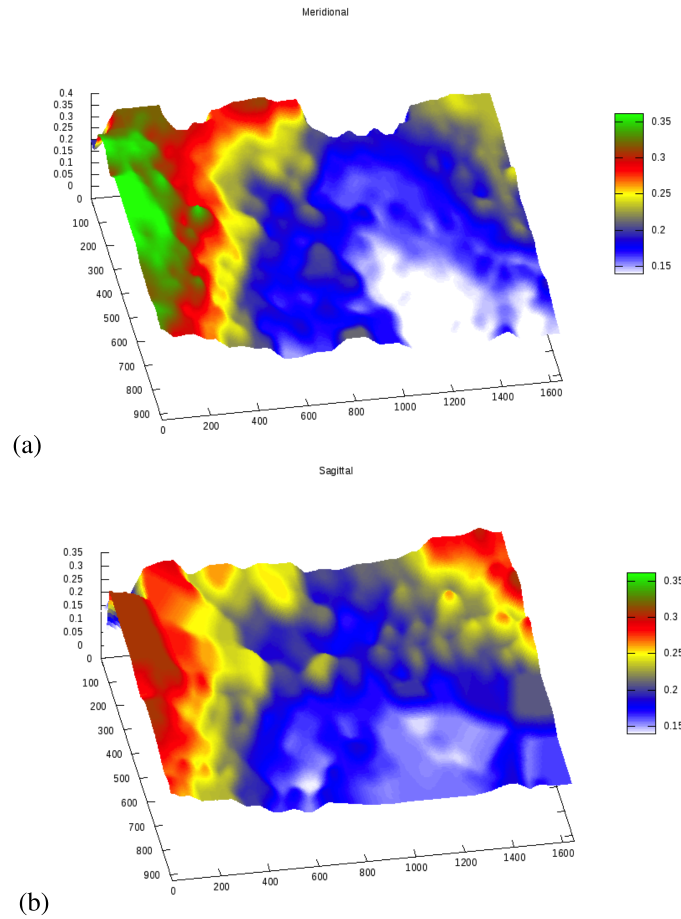

Fig. 11. 3D representation of MTF values for camera dome and lens system

V. PRELIMINARY FIELD RESULTS

The system has been deployed in the northern Sea of Cortez in Baja California, Mexico, to monitor the vaquita porpoise. Thus far, no vaquita detections have been made (either visual or acoustic). Figure 12 shows an example image from October 2015, which shows a small fish. However, this fish was detected by pure happenstance during a test of the camera system where the cameras were set to record at regular intervals, and did not trigger the camera via the acoustic system. Due to the scarcity of the vaquita, we anticipate that the system will be deployed for several months before a visual detection is made.

Other system tests have been performed in La Jolla Cove, California (Figure 13). Since our original use case was detecting the vaquita porpoise, which is exceedingly rare and has never been photographed underwater before, more work is needed to quantify the visual detection rate and compare with the acoustic detection rate using a more common species.

In order to make a claim about the efficacy of visual detections, we plan to deploy the system off the coast of San Diego to monitor bottlenose and common dolphins. Since these species are sighted much more frequently than vaquita, we will conduct a visual census and compare our observations to those captured by the system, in order to quantify the rate at which the system captures visual data when a species is nearby.

Fig. 12. Fish photo captured by camera in Sea of Cortez, Baja California. Note that this fish was not vocalizing and thus did not trigger the camera, but was an accidental photo captured by happenstance.

VI. APPLICATIONS

A. *Generalization to Other Species*

Although the system concept was originally developed to study the vaquita porpoise, we intend this to be a species-

Fig. 13. Frame grab from a video taken by the system, showing a pair of Garibaldi fish in La Jolla Cove, California.

invariant system for marine monitoring. Current work focuses on generalizing the system to detect Type 2 echolocation clicks, audible vocalizations, and pings from acoustic tags. However, in some respects the echolocation clicks serve as a better triggering mechanism for the purposes of collecting visual data because high-frequency signals attenuate more quickly underwater than low-frequency, audible vocalizations. For example, the low-frequency song of humpback whales can be recorded kilometers away, while the clicks of vaquita have only been recorded a maximum of approximately 200 meters away. For the purposes of visual detection, it is useless to start the cameras recording if a specimen is several kilometers away from where the system is moored, since the likelihood of the animal swimming within visual range of the cameras from that distance is small. If a specimen is detected 50 meters away, that likelihood is much higher. However, even if a specimen is too far away to be recorded by the cameras, the system still provides useful audio data. For this reason, we still intend to generalize the system to record other species even if their vocalizations are at a lower frequency.

B. Sensor Integration and Extensibility

The system is intended to be an extensible platform for complete ocean monitoring. Future work involves integrating other sensors in addition to the optical sensor payload, such as CTD, O_2, or pH sensors. These sensors can provide a more complete picture of the characteristics of marine habitats. Additionally, the onboard Intel Edison computer allows for implementation of more complex data processing algorithms, signal processing, and other user programmability.

C. Virtual Reality for Conservation Outreach

One interesting application of our camera system is in the area of Virtual Reality. The videos from each of the six cameras can be post-processed into a spherical video, which can then be displayed on a Virtual Reality headset such as the Samsung GearVR or the Oculus Rift. These visualizations can be used to effectively communicate the importance of conservation and ocean science to the general public in an engaging and immersive way. Our underwater camera system is currently being deployed in the Arctic to create spherical visualizations of underwater Arctic environments.

D. Machine Learning for Streaming Classification

Another potential application is using the onboard computational power to do streaming species classification, recognition, and detection, using the acoustic data. We are currently investigating machine learning techniques to do online detection and classification of acoustic features in order to improve the accuracy of the triggering system.

VII. Conclusion

This work described the design of an extensible, acoustically-triggered underwater camera system for marine species monitoring. This system is low-cost, in order to remain accessible to researchers in the fields of ecology and conservation. It is also extensible, allowing for simple integration of a variety of other sensors, such as CTD, O_2, or pH, and programmable, allowing for complex on-board data processing. Although the system was originally developed to study the vaquita porpoise, the system is being generalized to monitor a range of marine species, including other cetaceans, fish that emit vocalizations, and species tagged with acoustic pingers. Future field deployments and studies of different species will allow us to quantify the detection rate of the system across a range of species. This work provides a novel solution for coupling visual and acoustic data and thus facilitating autonomous behavioral monitoring of at-risk marine populations.

Acknowledgment

This work was supported by the National Science Foundation under REU Award #1560162, the National Geographic Society under Young Explorers Grant #C292-15, and the Association of Zoos and Aquariums under Grant #14-1217. Additional support was provided by the UC San Diego Frontiers of Innovation Scholars Program and The Explorers Club Exploration Fund—Mamont Scholars Program.

References

[1] J. R. Clemmons and R. Buchholz, *Behavioral Approaches to Conservation in the Wild*, Cambridge, United Kingdom: Cambridge University Press, 1997.

[2] F. Cazenave *et al.*, "SeeStar: A low-cost, modular and open-source camera system for subsea observations," in *IEEE OCEANS '14*, St. John's, NL, 2014, pp. 1-7.

[3] A. D. Sherman and K. L. Smith Jr., "Deep-sea benthic boundary layer communities and food supply: A long-term monitoring strategy," *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 56, no. 19-20, pp. 1754-1762, Sept. 2009.

[4] M. Doray *et al.*, "Joint use of echosounding, fishing, and video techniques to assess the structure of fish aggregations around moored Fish Aggregating Devices in Martinique (Lesser Antilles)," *Aquat. Living Resour.*, vol. 20, no. 4, pp. 357-366, Jan. 2008.

[5] R. L. Driscoll *et al.*, "Species identification in seamount fish aggregations using moored underwater video," *ICES J. Mar. Sci.*, vol. 69, no. 4, pp. 648-659, Feb. 2012.
[6] A. A. Myrberg Jr. *et al.*, "Shark attraction using a video-acoustic system," *Marine Biol.*, vol. 2, no. 3, pp. 264-276, Mar. 1969.
[7] 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*, vol. 61, no. 2, pp. 236-252, Feb. 2010.
[8] D. Mallet and D. Pelletier, "Underwater video techniques for observing coastal marine biodiversity: A review of sixty years of publications (1952-2012)," *Fisheries Research*, vol. 154, pp. 44-62, Jun. 2014.
[9] J. Assis *et al.*, "Performing fish counts with a wide-angle camera, a promising approach reducing divers' limitations," *J. Experimental Marine. Biol. and Ecol.*, vol. 445, pp. 93-98, Jul. 2013.
[10] M. C. Benfield *et al.*, "ZOOVIS: a high resolution digital camera system for quantifying zooplankton abundance and environmental data," in *ASLO Aquatic Sciences Meeting*, Albuquerque, NM, Feb. 12-17, 2001.
[11] "Report of the fifth meeting of the 'Comit Internacional para la Recuperacin de la Vaquita' (CIRVA-5)," Ensenada, B.C., Mexico, Jul. 8-10, 2014.
[12] D. R. Ketten, "The Marine Mammal Ear: Specializations for Aquatic Audition and Echolocation," in *The Evolutionary Biology of Hearing*, New York: Springer, 1992, pp. 717-750.
[13] G. K. Silber, "Acoustic signals of the Vaquita (*Phocoena sinus*)," *Aquatic Mammals*, vol. 17, no. 3, pp. 130-133, 1991.
[14] M. Lehaitre *et al.*, "Biofouling and underwater measurements," in *Real-Time Coastal Observation Systems for Marine Ecosystem Dynamics and Harmful Algal Blooms: Theory, instrumentation, and modelling*, Paris: Unesco Publishing, 2008, pp. 463-493.
[15] S. M. Wiggins and J. A. Hildebrand, "High-frequency Acoustic Recording Package (HARP) for broad-band, long-term marine mammal monitoring," in *2007 Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies*, Tokyo, Japan, Apr. 17-20, 2007.
[16] CR3 Hydrophone Specifications: http://www.cetaceanresearch.com/hydrophones/cr3-hydrophone/
[17] MTFMapper: https://sourceforge.net/p/mtfmapper/home/Home/
[18] Imatest SFR: http://www.imatest.com/support/modules/sfr/