Toward Remote Species Identification

Julia K. Parrish
University of Washington • Seattle, Washington USA

Introduction

Census of Marine Life (CoML) is an exploration of biomass and biodiversity. The Remote Species Identification (RSID) workshop concentrated on the latter category by exploring potential technologies to detect and identify marine species without capturing the organism (Table 1). Why is species identification important? Identifying new organisms (assuming they are related to known forms) gives us information about likely evolution, ecology, and physiology. If mapped, (e.g. species distribution) species identification allows us to track range alteration, introduction and extinction, the currency of biodiversity change/loss. If quantified, (e.g. species abundance) identification allows us to track biomass-biodiversity relationships which change as a result of both natural and anthropogenic forces. With regard to marine systems, RSID is useful because large tracts of ocean are inadequately sampled, because it is technically too difficult and/or financially impossible to do so.

Acoustics, optics, chemical ecology, and molecular biology experts were assembled for two days to explore the limits of their own disciplines, and then were tasked with marrying those technologies into an integrated, multi-sensor package – the Super-Predator. Instead of using sensors independently, perhaps we can take a lesson from biology and manufacture an integrated sensory system modeled after organisms which depend on species identification for survival. Super-Predators are machines or networks of machines which remotely detect and identify biological organisms, as well as a range of ancillary information which might include distribution and abundance, size, age, and health, individual identity, and environmental quality. Unlike biological predators, Super-Predators can possess a wide dynamic range of all relevant sensory modalities including those not currently available in the biological world (e.g. DNA fingerprinting), as well as be able to change the local physical environment to increase sensor abilities (e.g. light up a dark environment). During this workshop, three approaches to the Super-Predator concept were developed: a miniaturized, highly adapt-

Why is species identification important?

able system (S5), a platform-based neural net system (O3), and a set of example sampling strategies.

Super Sensor System for Survey & Synthesis – Rather than a static assemblage of sensors, S5 is a modular system designed to take advantage of integrated, miniaturized sensors. In this way, S5 can evolve from one form to another as conditions and human ingenuity warrant. Thus, deployment strategy is not an issue because S5 can be reconfigured to fit the needs of the problem/system. For instance, S5 can be an ambush (e.g. mooring) or pursuit (e.g. Remotely Operated Vehicle, Autonomous Underwater Vehicle) predator. In pursuit mode, S5 communicates with ship-board low frequency sonar which directs the initial pursuit path. As an AUV, S5 may dock to download data and repower (e.g. REMUS). In either predator mode, S5 may also communicate with passive listening arrays (e.g. SOSUS). As new sensor systems come online, S5 can evolve to incorporate them, either as add-ons or as replacements for aging technologies.

Open Ocean Observatory – O3 is an RSID stepping stone, based on *Deep Sea Observatory* a proposal to convert an oil drilling platform into a mobile marine laboratory. During the initial deployment phase (approximately 2 years per site), O3 acts as a neural net with a central brainstem (the platform) which directs the actions and flight-paths of series of more mobile nodes (AUV, ROV, ABE – autonomous benthic explorer – ships with towed arrays) as well as the sensing systems on any moored arrays with fixed, low frequency acoustics which localize centers of biological activity. Remote and manned vehicles equipped to collect novel biota are deployed. These systems are also equipped with a set of integrated sensors (e.g. BIOMAPER II) such that remotely sensed libraries can be developed and groundtruthed by scientists onboard O3. Because O3 is manned, scientists can also direct the deployment of remote sensing technologies to resample areas of interest. Because O3 is mobile, it can be deployed to remote areas where previous sampling has been inadequate. Once area and technology-specific libraries have been developed and tested, O3 will move to a new loca-

tion, leaving behind an RSID mooring for time-series, multi-sensor data.

Strategy Concepts – A Super-Predator must detect, identify, and enumerate. Detection and identification requires analysis and interpretation of one or more signals. Tools operate on different scales, and are also useful for different aspects of this three-pronged requirement. Given the wide range of size and mobility of marine organisms to be censused, as well as the challenge of designing a single Super-Predator with all available sensors, we envision Super-Predators as sampling strategy concepts rather than machines per se. We offer two example strategies:

1. *Cascade* – Use a biochemical attractant to "seed" or catalyze a series of predator-prey interactions. Essentially, create a hotspot or ephemeral community within a defined spatial framework which can be sampled definitively. Both stationary and mobile sensor packages can be deployed to identify and count the organisms assembled.

2. *Dynamic* – Use coarse-grained sensors (e.g. acoustics) to identify areas of biological activity (hotspots). Further sampling can be targeted in these areas. "Robotuna" or other biosensors can be used to track down individual species within the hotspot.

Conclusions

Participants in the RSID Workshop concluded that remote species identification is a possibility which should be explored, but that there are both guidelines and constraints. These fall into three basic categories: program relevance, identification issues, and design approach.

Relevance – Just enumerating species (i.e. a catalog of species richness) is not a worthy objective. Distribution and abundance is a must. Ecological (e.g. migration, life history, trophic dynamics, etc.) and environmental measures (e.g. physical habitat, water properties, pollution, etc.) are also important data. Because the oceans are vast, knowing something about how the physical environment is structured may help us narrow our search effort for biology over a uniform sampling strategy.

All sensory systems rely on a reference library for matching/identification. Given a sufficient library, pattern recognition systems – whether acoustic, optical, chemical, or DNA-based – can be trained to detect species remotely. However, humans are needed to create the libraries necessary to train multi-sensor systems and may also be needed to perform the ultimate task of decision-making relative to truly new species. This is because the concept of a species is not operationally defined, but rather a subjective decision on the part of the taxonomist. This assertion has two important caveats. First, taxon-specific experts are needed now and into the future. Without this knowledge base, it will

not be possible to train Super-Predators, regardless of how clever their detection and identification systems are. Second, the necessity to ultimately have the specimen in hand does not preclude remote species identification. Super-Predators as outlined here may be the only way we can adequately sample habitats such as the deep sea or open ocean within the next few decades.

Identification – There is a difference between designing a system to enumerate the known and one that also begins to catalog the unknowns. Both things are important. There is also a difference between unknown and undiscovered. Undiscovered things inhabit places we have undersampled (e.g. deep sea), are destroyed by traditional sampling techniques (e.g. gelatinous zooplankton in nets), or are very small (e.g. viruses and bacteria). Cryptic species (the intersection of known and unknown) result from lumping sister species into a single named taxon. In general, we should concentrate on what we don't know, including undiscovered species and undersampled habitats. We should also consider concentrated effort in biological hotspots.

Design Approach – Acoustics and optics have dominated species identification and enumeration to date. It is not clear that acoustics alone is sufficient to correctly identify species, especially as the number of species/organisms increases and/or the number of previously undiscovered species increases. Super-Predators can use large-scale technologies (e.g. low frequency acoustics) to direct the use of other more fine-scale modalities (e.g. optics). All technologies are evolving rapidly, making it difficult to predict what will be available within the next 10 years. The positive side of this uncertainty is that with some redirected effort, currently under-utilized technologies, particularly those centered on biochemistry, can come "on line" relatively quickly. For instance, there may be potential to develop biosensors - blends of organism sensory structure and machine amplifier - to detect chemically-based taxon-specific signals.

All technologies are evolving rapidly, making it difficult to predict what will be available within the next 10 years.

Proximity is a requirement for most types of identification of species (whether remotely or via capture). Therefore, Super-Predators must be stealthy - they should not provoke unwanted responses, either repulsion or attraction, or the use of attractants (e.g. light, sound, chemicals) should be seriously considered.

Virus-to-whales organism scale variability is a serious issue for Super-Predator design. All organisms cannot be addressed by a single, even multi-sensor system. However, a network of Super-Predators may be able to overcome the scale problem. Super-Predators should be modular to deal with the issues of deployment, target organism size, and evolving/new technologies. To function most efficiently, Super-Predator design relies on sensory integration and signal processing for internal groundtruthing. These issues need to be explored more fully.

TECHNOLOGY	ORGANISM	HABITAT	PLATFORM	AVAILABILITY
Passive Acoustics (20 Hz - 30 KHz)	marine mammals some fish & invertebrates	potentially all	fixed - mobile	array - now
Active Acoustics[1] (38 KHz - 1MHz) (1 - 200KHz)	zooplankton fish	water column water column	fixed - mobile	off-the-shelf
Photography standard stereo	all marine mammals to zooplankton	nekton - benthic nekton - benthic	all all	now
Video standard stereo	large plankton, small fish large plankton, small fish	nekton - benthic nekton - benthic	all all	now
VPR[2]	all plankton	plankton	all mobile	now
LLS[3]	marine mammals to zooplankton	benthic	all mobile	now - future
Holography	all plankton, small fish	plankton	all mobile	now - soon
Bioluminescence non-imaging imaging	plankton, small invertebrates plankton, small invertebrates	nekton - plankton nekton - plankton	all all	now - soon
Fluorescence imaging non-imaging	all plankton	plankton - benthic plankton - benthic	ROV, AUV, sub diver, ROV	now - soon
LIDAR[4]	marine mammals, fish	nekton - benthic	air	soon - future
OPR[5]	all plankton	plankton	all mobile	soon
Signature Molecules	all	benthic	all	1-5 yrs
Chemical Attractants	carnivores, omnivores, reproductive	all	all	some now
Biosensors	all	all	all	5-10 yrs
Surface Moieties	small plankton	plankton	all	5-10 yrs
DNA-based (hybridization)	all	all	all	some now applications: 3-5 yrs

1. Can identify type but not usually species. 2. VPR = Video Plankton Recorder. 3. LLS = Laser Line Scanner. 4. LIDAR = Light Detection and Ranging System. 5. OPR = Optical Plankton Recorder