REPORT ON THE CURRENT STATUS AND FUTURE OF BEHAVIORAL RESPONSE RESEARCH

Report prepared by:
Dr. Catriona M. Harris
Dr. Len Thomas
University of St Andrews

Workshop reviewers:
Professor Vincent Janik, University of St Andrews
Dr. Hans Slabbekoorn, University of Leiden
Professor Douglas Wartzok, Florida International University

Sponsors:
US Office of Naval Research (ONR)
US Living Marine Resources (LMR)
US National Oceanic and Atmospheric Administration (NOAA)

Executive Summary

A review of the status and future of research into behavioral responses of marine mammals to naval sonar exposure was undertaken to evaluate the return on investment of current US Navy funded programs, identify the data needs and the contributions of current research programs to meeting data needs, and determine the ability to meet outstanding data needs given the current state of technology.

As part of this review, a workshop was held from 21-22 April 2015 in Monterey, California. Workshop attendees were key representatives of Navy-funded behavioral response studies, as well as three external reviewers who were selected because of their expertise in animal behavior and behavioral responses to anthropogenic stimuli in the aquatic and terrestrial environments. Prior to the workshop, a questionnaire was circulated to canvass the opinions of members of the scientific community (primarily workshop attendees exclusive of external reviewers) on each of the research approaches taken to address this topic. The workshop was then structured around the questionnaire and responses received, via a series of discussion sessions. Afterwards, each research approach was evaluated independently by the external reviewers. This report presents a synthesis of the evaluations and recommendations of the external reviewers on current and future behavioral response research relevant to naval sonar.

All reviewers agreed that excellent progress has been made on this topic and that each of the research approaches has contributed to our understanding of cetacean responses to naval sonar. The report includes specific comments and recommendations of the reviewers relevant to each approach, but also includes suggestions for priority species and a comprehensive list of recommendations for the future of BRS research in general (Tables 1 and 2). In summary it was recommended that BRS research be continued and extended to increase sample sizes and experimental replication, and temporal duration and spatial scale including more research in areas where the animals are presumably more naïve than on the naval ranges. It was noted that future investigations would benefit from combining experimentation and observation to enable linkage of short-term behavioral response to long-term fitness consequences of repeated exposure. Beaked whales were the species group ranked highest in terms of research priority. The importance of baseline studies and longer-term monitoring of animals before and after exposure is emphasized throughout.

Contents

Introduction

Many marine mammals rely on sound for foraging, maintaining group cohesion, navigation, finding mates and avoiding predators. Hence, they may be profoundly affected by the introduction of anthropogenic noise into the marine environment. Examples of potentially harmful noise sources include vessel noise and active acoustic devices such as naval sonar or airguns used for seismic prospecting (Richardson et al. 1995, DeRuiter 2010). Potential adverse effects of those sounds include reduction or cessation of feeding (Miller et al. 2009, Goldbogen et al. 2013), strong avoidance responses (Tyack et al. 2011, DeRuiter et al. 2013, Miller et al. 2014), and stranding (D'Amico et al. 2009). Ultimately we would like to understand the overall fitness consequences of sound exposure to individuals and populations and thus behavioral responses are not the endpoint of study; rather they can be thought of as modifiers of potential risk in that they increase or decrease the exposure level received by the individual, with the associated physiological risks, and they can be advantageous or (likely more often) disadvantageous in terms of ecological opportunities. For example, strong avoidance responses may exclude animals from important habitats, alter opportunities to find and consume prey, modify predation risk, or result in separation of dependent offspring and mother (Miller et al., 2012). Therefore, it is important to understand the probability of different types of behavioral responses occurring under particular sound exposure conditions and what the severity and subsequent consequences of a response might be.

Within the United States there is a requirement for federal agencies to estimate behavioral impacts to marine mammals that result from noise-generating activities. This is achieved through the use of risk functions, which predict the probability of a behavioral response as a function of the sound pressure level (SPL; dB re 1 m Pa) received by the animal (i.e., the received level), and density estimates for each species that is likely to be present in the affected area. The risk functions that are currently used in U.S. Navy environmental compliance documents are data poor. Consequently the Navy has invested in a number of research programs with the aim of providing data to support environmental compliance.

Behavioral Response Study (BRS) is the term used to encompass the research efforts that have the common objective of determining the relationship between the dose of a stressor (which can be represented by many different metrics) and behavioral response. Here we refer specifically to the study of behavioral responses of marine mammals exposed to Naval sonar sounds, in particular hull mounted mid-frequency active (MFA) sonar. A number of different approaches have been adopted in carrying out this research and some broad distinctions can be made. The first distinction is between studies that rely on a formal, pre-determined experimental design and those that don't. We will refer to those with an experimental design component as controlled exposure experiments (CEEs) and those without as observational studies. Within the CEE approach there are those that have been carried out in captive facilities using trained animals and there are those that have been carried out in the wild setting using free-ranging animals. Within the observational approach there are those that have employed passive acoustic monitoring (PAM) methods and those that have used medium- to long-term tagging methods (e.g., satellite tagging). Within each of these categories there are many more distinctions to be made, for example there are differences in the sources used within CEEs (simulated sonar versus real sonar), and many differences in protocols relating to signal type, length of exposure, source position relative to focal animal, to name just a few. The choice relating to each of these parameters relates to the question(s) being addressed by the study.

This report is the product of a review of U.S. Navy investment in marine mammal BRS, in particular the research approaches mentioned above. The goal of the review was to:

1. assess the current state of knowledge of the topic in order to evaluate the return on investment of current research programs,
2. identify the data needs and the contributions of current research programs and methodological approaches to meeting data needs, and the ability to meet outstanding data needs given the current state of technology.
3. support the long-term goal of understanding research progress and needs for future research by providing a comprehensive critique of the state of the art in marine mammal behavioral response research.
4. formulate recommendations for future behavioral response research.

The primary component of the review was a workshop held 21-22 April 2015. The workshop participants were key representatives of each of the research programs (Appendix 3) plus three external reviewers (Appendix 4) who represented expertise in animal behavior and behavioral responses to anthropogenic stimuli in the aquatic and terrestrial environments. Prior to the workshop a questionnaire (Appendix 1) was circulated to canvass the opinions of members of the scientific community (primarily workshop participants exclusive of the external reviewers) with respect to each of the different research approaches. The structure of the questionnaire and the collated responses formed the structure of the workshop, which consisted of a series of discussion sessions (Appendix 2). After the workshop, the external reviewers each provided an independent evaluation of the research conducted to date and recommendations for future research efforts, using as source material supporting literature, pre-workshop presentations provided by a representative of each research approach, the questionnaire responses and the content of the workshop discussion sessions. The three independent reviews can be found in Appendix 5. The role of the authors of this report has been to synthesize the three independent reviews and ensure all views are represented in the main body of the report, which contains a synthesis of the evaluations and recommendations. There was a high degree of consensus among reviewers (after discussion and editing of the draft document) in their recommendations; we note in the text where all reviewers did not agree.

The report is structured as follows. The next section contains an overview of the progress made to date as a result of the U.S. Navy investment in the BRS research programs. There follows a brief overview of each research approach in turn, with comments and recommendations made by the reviewers that are relevant to each approach. We briefly comment on the utility of predator playback studies. We then provide a comprehensive list of key research topics identified and recommended by the reviewers as high priority for future research effort. These span all research approaches and full rationale for each recommendation is provided. Tables 1 and 2 capture all the recommendations contained within the report and provide an indication of priority as scored by the reviewers. Each recommendation in Table 2 is linked to the relevant text in the report using superscript numbers. The report concludes with some general comments made by the reviewers about the BRS research program.

Overview of Progress

There has been tremendous progress in a number of areas that provide greater understanding of cetacean responses to naval sonar and the potential consequences. The Navy program has funded many excellent studies on this topic. A number of different approaches have been supported and each has yielded useful data and valuable insights. Each approach has strengths and limitations, and several of the available methods have clear potential to be complementary. Overall the program has to be complimented on a tremendously successful effort to further our understanding of how cetaceans react to sonar. The following review consists of a series of recommendations to continue its success and address the questions that have emerged from the data that have been collected so far.

The development of the DTAG is perhaps one of the most influential developments of the Navy research program. The usefulness of this tool should not be underestimated, not only for sonar studies but for marine mammalogy in general. It has allowed detailed measurements on underwater behavior to be collected during controlled experiments on individuals in the wild, and this is and will remain one of the most useful ways of studying responses to noise. The experimental approach that can be used when tagging animals allows us to compare the reaction to treatment stimuli in comparison to control stimuli. This comparison provides a scientifically strong assessment, which cannot be achieved by comparing reactions to sound sources with baseline behavior.

Tantalizing potential answers have been proposed for several of the initially most troublesome issues. Initial studies on beaked whales have shown interesting reactions to sonar, either using passive acoustic monitoring (e.g., in the AUTEC range) or DTAGs to record responses on the animal. Studies on more accessible species have contributed a tremendous amount to our understanding of cetacean reactions to sound. Species like pilot whales, blue whales, and killer whales were more accessible and therefore studies resulted in larger sample sizes. Passive acoustic monitoring studies supplementing DTAG work showed interesting reactions but are limited by their dependence on vocally active animals. Captive work has been used to test tolerance of trained animals in training contexts, but there are concerns about how these data translate to wild, untrained animals. It is unknown whether wild animals will be more or less sensitive than captive animals.

The need now is to extend this work through increased sample sizes and study replication, extended temporal duration and expanded spatial scale including more research in areas where the animals are presumably more naïve than on the naval ranges. Future investigations are likely best to be a combination of experimentation and observation that enable linkage of short-term behavioral response to long-term fitness consequences of repeated exposure. An explicit goal and requirement for the interpretation of results from these studies is to understand the acoustic world of animals, including the natural use of sounds for communication and orientation.

Critical Review of Research Approaches

In this section we address each of following research approaches in turn:

1. Controlled exposure experiments (CEEs) on captive animals using simulated sources of Navy sonar
2. Controlled exposure experiments (CEEs) on free-ranging animals using simulated Navy sources or real, but scaled, Navy sources on research vessels
3. Controlled exposure experiments (CEEs) on free-ranging animals using real Navy sources deployed by Navy vessels
4. Observational studies (primarily using tags, PAM and visual observation) in relation to exposure to real Navy sources and Navy vessels
5. Predator playback studies

Following a brief introduction to each approach we provide a synthesis of the reviewers' comments and then a list of the key recommendations for future research that are relevant to that approach.

Controlled exposure experiments (CEEs) on captive animals using simulated sources of Navy sonar.

Introduction

Controlled exposure experiments have been conducted on a small number of species within captive facilities. Until recently studies on captive animals primarily focused on understanding species' hearing abilities and relating sound exposure to hearing damage. However, in the last few years there have been studies that have focused on behavioral responses of trained individuals in response to exposure to different sound stimuli. Such studies using simulated naval sonar signals have been conducted on California sea lions (Houser et al. 2012, 2013), bottlenose dolphins (Houser et al. 2012, 2013) and harbor porpoise (Kastelein et al. 2011, 2013). In these studies animals were trained to perform a certain task. They then performed this task during a number of control and exposure trials whilst a number of metrics were collected relating to their behavior and exposure levels (dose received). These studies have allowed the construction of probabilistic dose-response relationships for these species in settings with a high degree of experimental control and the ability to take many detailed measurements. These studies have highlighted species differences in thresholds for response and response severity, but also differing inter-species variability and the potential role of intrinsic factors (e.g., age and sex). One major constraint of this approach is the use of captive, trained individuals and our lack of understanding of how their responses can be translated into the wild context. We know from other studies that environmental and social contexts probably play a crucial role in the probability and severity of responses, and that these factors cannot be explored in the captive setting.

Reviewer Comments

Captive studies have complemented free-ranging animal studies. Harbor porpoise have occasionally stranded and in general show the greatest sensitivity to acoustic disturbance. Captive studies have provided dose-response data for these sensitive species that are too small to tag in the wild with current data loggers. A lot of good data has been obtained from mainly one captive porpoise but so far nothing from wild animals of that species (Kastelein et al. 2011, 2013). For other species (bottlenose dolphins and California sea lions) Houser et al. (2012, 2013) have been able to conduct studies with greater sample sizes and obtain dose-response curves with multiple animals exposed to a given signal level. These studies have demonstrated that the observed variability in responsiveness of animals in field studies is to be expected, as different shaped dose-response curves (symmetrical and asymmetrical), different evidence of habituation, and different age-specific sensitivities across different species have all been observed in captivity.

Captive studies are often criticized for not being representative of behavior in the wild. However, they have a role in testing directions of responses that can then be further studied in wild animals. For example, general aversiveness of parameter combinations can be tested in captive animals. The immediate reflex reaction to such novel stimuli should be similar to those in wild animals and will allow investigators to decide what to test further in the wild. Captive studies could offer a much more efficient way to narrow down the parameters of interest.

Captive studies can be useful to get significant insights, especially if we accept that, in this setting, dose-response relationships and absolute threshold criteria are not the single-most-important target of study. Behavioral effects occur within the audible range of a species determined by detection thresholds at the low-end and more or less by physical damage thresholds at the high-end. There is growing awareness that it is critical to study variation within these broad limits and to gain insight into the many factors that determine whether or not a behavioral response occurs and whether or not there is a detrimental effect. Several factors can be investigated best with optimal exposure control and knowledge about individual characteristics and experience. Consequently, captive studies could for example contribute to our understanding of the effects of temporal patterning, the potential for habituation and sensitization and the variability associated with individual personality.

Personality refers to variation among individuals of the same species reflected in different response tendencies that are consistent across time and space. Variability in personality may mask or bias our interpretation of general behavioral response patterns, especially in the case of small sample sizes. One reviewer suggested that, for species with some group social structure, typical response patterns in the field may be determined and assessed at group level and therefore individual variation will average out and be less important. He also noted that it may be the other way around in that group responses are determined by the most easily frightened and most responsive individual. If this is the case, it would elevate individual personality to top priority in predicting response thresholds at the group level. Another reviewer proposed an additional hypothesis whereby in areas where there is some history of exposure, more sensitive animals may have left the area, and so prior exposure reduces the variability associated with personality.

Recommendations

At one point it was thought that CEEs would be able to elucidate the characteristics of the sonar signal that trigger responses in marine mammals. Research at this stage seems to indicate that this is not a likely outcome of the CEE experiments on free-ranging animals. [1]Determining the particular characteristics of the sonar signal that are the most critical for generating a response would be a suitable subject for captive research but there would still be major questions regarding generalization to other species. (Recall that superscript numbers in the text link recommendations made here to the summary in Table 2.) Species amenable to captive research studies are not the species particularly sensitive in the wild.

[1]Captive choice studies can help to differentiate between different proposed response mechanisms. These are studies in which animals are presented with different sounds in different parts or pools of their enclosure to observe where they choose to spend more time indicating a greater tolerance for the associated signal.

Captive work has been used to test tolerance and cooperation of trained animals in training contexts, but there are doubts whether these data represent tolerance in wild, untrained animals. [2]More could be done here by exposing captive animals unexpectedly to sound and by, again, using [1]choice experiments to determine preferences and aversion to acoustic stimuli.

[2]The captive setting could be used to investigate potentially detrimental effects on performance during natural tasks that are critical for determining vital rates. There is, for example, growing evidence in a wide variety of taxa that elevated sound levels negatively affect performance in tasks that do not even need to have an auditory component. Responses to visual stimuli can be delayed or scores to cognitive tests can be lower due to noise-induced attention shifts or neural processing efficiency. Captive animals may for example be used to test whether marine mammals that remain in an area and that do not show any conspicuous behavioral response may still be negatively affected by audible sonar if exposure has the potential to lower foraging efficiency.

[3]New technologies could also be tested on captive animals to save time and money when deploying them in the wild. For example, more advanced physiological tags could be tested on captive animals.

Controlled exposure experiments (CEEs) on free-ranging animals using simulated Navy sources or real but scaled Navy sources on research vessels

Introduction

In a free-ranging CEE the focal species is selected based upon research need and the focal animal is the individual which becomes the focus of the study, either because it was successfully tagged with an animal-borne tag or because of the ability to follow it visually. Its behavior is then monitored using visual observations, passive acoustics, animal-borne tags or a combination of these. After pre-exposure observations, the tagged whale is exposed to a stimulus, such as a naval sonar sound or control sound, and its response is monitored. In many of the experiments the dose of sound increases over the duration of the exposure and therefore they can be thought of as dose-escalation studies. Various measurements are recorded before, during and after exposure, including location (in 3D) through time, vocal behavior, underwater orientation, and behavior observed at the surface.

To date most of the free-ranging CEEs have been carried out using scaled or simulated sonar sources, which has meant the research vessel has been relatively close to the focal animal to achieve the desired received sound levels. The effect of the research activities on responsiveness remains unknown, as does the relationship between responses to simulated and real naval sources.

A number of project teams have undertaken these types of studies in different geographical locations and have each focused on different species, although beaked whale species have been a common focus across all studies (Miller et al. 2011, 2015, Tyack et al. 2011, Kvadsheim et al., 2011, 2012, 2014, Southall et al. 2012). These studies have provided experimental evidence for avoidance of sonar by certain species, at lower levels than predicted for some species. They have highlighted high levels of intra- and inter-species variability in responsiveness and severity of response, and have started investigating the role of context in understanding this variability. Dose-response functions have been produced for some species (Miller et al. 2014, Antunes et al. 2014, Harris et al. 2015).

Limitations of this approach include the experimental context, the short exposure and observation periods, and technological constraints of tagging certain species groups (e.g. small delphinids). As with any study, the transferability of the results from the studied (and normally tagged) individuals to the wider population is unknown.

Reviewer Comments

These studies gain in realism relative to the captive studies at the expense of some control over experimental conditions and insight into animal characteristics and individual and group histories. However, there was a general positive attitude at the workshop to the accomplishments using this approach. The careful stepwise approach to the CEEs has built the confidence of researchers in, and mollified the concerns of most of the sceptics to, the CEE approach.

The DTAG has allowed detailed measurements on underwater behavior to be collected during controlled experiments on individuals in the wild, and this is and will remain one of the most useful ways of studying responses to noise. The experimental approach that can be used when tagging animals allows us to compare the reaction to treatment stimuli in comparison to control stimuli. This comparison provides a scientifically strong assessment, which cannot be achieved by comparing reactions to sound sources with baseline behavior.

Early CEE studies showed that the response of tagged beaked whales was to break off foraging dives, ascend to a mid-depth, pause, and then ascend slowly while increasing the distance between the whale and the source. This was an important finding because it disabused a notion in the literature that the negative sequelae in beaked whale responses to sonar were the result of decompression sickness (DCS) triggered by a rapid ascent.

Reviewing dive profiles with and without controlled exposures showed that in most cases the responses of the cetaceans to sonar would not lead to nitrogen tensions resulting in DCS. The theoretical work on compartmental distribution of nitrogen during a dive showed that in the animals where there was a possibility of DCS the critical factor was not their deep dives or ascents therefrom but rather an interplay between the time spent at moderate depths (30-200 m) during which nitrogen loading could occur compared to time spent in repetitive shallow dives (<30 m) where nitrogen could be safely flushed through alternating cycles of decompression and recompression.

The early work was designed to address directly the response of beaked whales to sonar because the stranding data indicated that these species were the most sensitive (Cuvier's beaked whale, *Ziphius cavirostris*, constitutes 69% of all mid-frequency active sonar strandings). It soon became apparent that the most sensitive species were also the most difficult to study. Hence the overall study design was re-thought and the focus broadened to embrace other species that were more amenable to tagging and study but that also were likely to be affected by sonar, just not in as dramatic a manner. A second important new direction was to redesign the sonar simulation source so that it could be deployed from smaller vessels at a variety of locations. The wisdom of both of these redesigns has become evident through the substantial increase in CEE subjects.

Along with range, direction of movement of the source can be a significant influence. Other studies of cetacean responses simply to vessel noise have shown greater response to vessels approaching than to those stationary or moving away. The 3S experiments showed that the animals changed direction to always move perpendicular to the vector of the ship's movement.

Recommendations

[4]In order to determine the extent to which the changes in behavior observed thus far are truly atypical, more baseline data are required. The baseline certainly needs to be extended around the time of CEE, both before and after. Longer term tag deployments would help with gathering more baseline data.

[17-20]A longer duration tag could also address concerns regarding research effects, particularly those associated with the tagging. An argument can be made that research effects would affect the control and experimental presentations equally but it is much better to remove those effects than to have to make the argument, particularly when the extinction coefficient of the research effects is unknown and is likely different for different species and different behavioral states within a species. The longer life tag would also allow more experimental presentations and an extended assessment of post-exposure behaviors.

The strength of the tagging and exposure studies is the potential to effect proper controls.   CEE-studies including tagging and subsequent experimental exposure need to conduct control trials in which every aspect of the experiment is the same, except for the sonar exposure. Treatment and control trials should be assigned randomly or in a pre-determined balanced way to avoid effects of order, time or day, season, region, etc. In that way, any impact of vessel presence and tagging can be controlled for. Special care should be taken to achieve such control trials where possible. A group or individual can also be its own control by using baseline data from after tagging and before exposure to compare with behavior during and after. However, baseline data are often very variable, making it hard to reveal any significant change in behavior at the onset and during the sonar exposure. Therefore, again the ability to achieve longer periods of tagging would help, and allow more selectivity in the periods that are used as baseline, really matching the particular environmental conditions and behavioral mode the animals were in at the moment of exposure.

The focus on the DTAG has perhaps decreased observational efforts and the group focused sampling methods for surface observations seem perhaps somewhat unsuitable to evaluate reactions at the level of detail that is required. [6]More efforts to expose animals while at, or near, the surface combined with observational methods that do not require tag attachment but follow individual animals could be used to increase sample size in a range of species, perhaps even beaked whales despite the difficulties of visually following these species. Exposures carried out when animals are actually at the surface will result in lower received levels and so exposures would be better timed when the animals are near the surface, for example shortly after the initiation of a dive.

[7]There is a lack of studies with two sources, or repeated exposure of the same individuals over several days to better understand cumulative effects. [8]Similarly, different source movements may lead to different reactions. The 3S project has applied such an approach, but did not compare it sufficiently to non-directed passes.

[9]The suggested pattern of species differences in individual variability requires further investigation under more conditions and in different areas on the same species. Furthermore, this theory would gain external power with data on more species. We also need to know what this means in terms of impact: is it more or less harmful to a population if many individuals are affected to a moderate extent or few individuals to a large extent? It will probably be necessary to find out whether variation in response tendencies between and within species is actually related to the detrimental impact of the disturbance. Some species may alter their behavior quickly to anthropogenic stimuli as well as to many natural stimuli throughout the day, and even a strong modification of behavior by sonar may not result in a significant

disturbance that would be measurable energetically. Other species may be less easily affected in changing their swimming, direction or foraging activity and for them a subtle behavioral change in response to sonar may actually have more serious consequences in terms of increased energy expenditure.

Controlled exposure experiments (CEEs) on free-ranging animals using real Navy sources deployed by Navy vessels

Introduction

The difference between this approach and the approach described above is the use of real Navy sources and vessels rather than simulated or scaled sources on research vessels. The overall approach still uses an experimental design but relies on coordination with Navy vessels to carry out the sonar exposure. The exposure is therefore more realistic in terms of source characteristics and source intensity. This allows for study of the relative contributions of received sound intensity and range-to-source in determining response. The primary limitations of this approach are the logistical challenges of coordinating animal tagging with the availability of Navy vessels, which results in low sample sizes. This approach was trialed in 2013 and 2014 in Southern California and, despite the small number of exposures, results are beginning to indicate possibly important differences between simulated and actual sonars.

Reviewer Comments

Several studies have indicated that both distance and received levels may both be important predictors of response. To date this conclusion has been based on few observations with other unknown confounding variables which further complicate the picture. However, some animals use the frequency-specific changes in a sound that arise during transmission as an indicator of the distance to the sound source; for example we know that birds can estimate distance due to the changing properties of sounds transmitted through their typical environment and it is possible that marine mammals do the same. Animals can also use their directional hearing to determine distance. Regardless of the mechanism for determining distance, the interaction between distance and received levels would not be at all surprising if the animals perceive the sonar as a potential predator. Prey typically titrate response against distance and do not flee with the first detection of a predator unless the predator is within range to threaten them. Understanding the relationship between received level and range is an essential issue for translating the CEE into values that can be used for regulatory purposes. It is unlikely that the dose-response functions derived with the simulated sources represent the dose-response functions with exposure to operational sonar where range is an important factor. It is believed that the best data for disentangling the relationship between distance and received level could come from CEE in conjunction with Navy ships, but only if the logistical constraints associated with these studies allow for sufficient replication and control. Unfortunately, given the difficulties of coordination with Navy ships for CEE much of these data will need to be derived from experiments with scaled sources as well as observation of tagged animals in less controlled encounters.

Recommendations

[10]CEEs with real naval vessels should be conducted to gather data to generate dose-response functions and to [11]disentangle the relationship between distance and received level. The challenge will be achieving reasonable sample size and distance control between animals and sound source at sea. In terms of experimental design in an ideal world, it would not be too difficult to tease these factors apart. To the extent that distance is perceived by frequency-dependent changes in the signal with distance, one could test response tendencies for a replicate set of sonar sounds that are recorded at two different distances and played back from the same distance at the same normalized level (Signal altered/Received level constant). One could also test response tendencies to a replicate set of sonar recordings that are played back at two different levels from the very same distance (Signal constant/Received level different). Each test would be strongest in a paired design if it were possible to test individuals or groups twice to get responses to two exposures. If so, the order obviously should be alternating and taken into account statistically. Replicate set here means that every pair of sounds played back to another individual or group ideally should be a unique recording. The replicate set of recordings should then be a sample that reflects the natural variation in sonar exposure conditions that we are interested in. In reality such an experimental design may be impractical in this context. The detailed tracks provided by DTAGs can be used to investigate whether the animals are using angular changes in direction to the source to determine distance from the source. The observation that killer whales consistently moved perpendicular to the path of the source vessel in the 3S experiment indicates an ability of killer whales to use directional hearing to maintain a heading always perpendicular to the track of the source vessel.

Observational studies (primarily using tags, PAM and visual observation) in relation to exposure to real Navy sources and Navy vessels

Introduction

Observational studies are those where there is no experimental design and the researcher does not control the sonar operations, and therefore they are very different from the CEE approaches described above. Within the observational approach there are a number of different methodologies that have been employed to monitor the presence, movement and behavior of marine mammals. These methods fall into the following categories: 1) A combined approach which uses both animal-borne tags and PAM; 2) medium- to long-term tag studies; 3) PAM studies combined with visual observations; 4) PAM studies where PAM is used to track animals; 5) PAM studies where PAM is used to gather detections only (with no attempt at localization); 6) population level studies, e.g., using photo-identification methods. Gathering data on the potential effect of sonar using these methods is possible because they are often employed within U.S. Navy training ranges or off-range but near regions of high sonar use.

The first method, where tagging and PAM are used concurrently, has not yet been carried out around a known exercise, although it is proposed. The idea would be to tag animals on a training range prior to an exercise and then collect data throughout the exercise on the animal and also from bottom-mounted hydrophones on the range. The main difference between this and the CEE using real naval sonar is the lack of experimental design.

Tagging studies in the observational context refers to the deployment of medium- to long-term tags, such as Argos satellite tags, on animals in the vicinity of naval training ranges (Schorr et al. 2014). The availability of good information on the characteristics of sonar exposures and the ability to accurately model received level at the animal depends partly on whether the animal was on the range during periods of sonar activity. If the animal is off the range then the acoustic data are less reliable.

PAM methods have been combined with visual observations to study a range of species, some of which are currently too difficult to tag. The visual observations have been used to record the behavior of individuals and groups over time when in the vicinity of the research platform. The concurrent deployment of static hydrophones has resulted in the incidental detection of sonar activity, allowing for the data to be examined for any evidence of behavioral disturbance (Henderson et al. 2014).

Methods have been developed that allow animals to be tracked using PAM and this is being done both on and off navy training ranges. However, to date there have been no studies that have looked at individual animal tracks in relation to sonar exposure.

Some PAM deployments can only produce animal detections and can't be used for generating animal tracks. These can be useful for long-term monitoring of animal presence and measures of relative animal density, but method development is required to link animal detections with sonar detections.

Finally another method for long-term population monitoring is the use of photo-identification to look for population level changes in areas with and without sonar activity. Such long-term studies can help to provide context and, as datasets become longer, then it is likely that behavioral shifts will become more detectable.

Each of these observational approaches to the study of behavioral response has its own strengths and weaknesses. One of the main strengths is the reality of the context as these studies are carried out within the full context of actual naval training and monitor behavior over more realistic time-scales than CEEs. Another strength of PAM and photo-identification studies is that, unlike tag-based studies, we can be sure that animals remain undisturbed by the measurement procedure. However, one of the main limitations is the lack of formal experimental design. One limitation particular to PAM is that it can only be used to study animals that make sound, and our ability to monitor vocal individuals or groups ceases when they go quiet.

Reviewer Comments

[13]A combination of methods as has been used in the AUTEC studies is often helpful to maximize our understanding. The AUTEC geographic setting is quite unique and offers many advantages for method integration, for example the geographical layout can aid interpretation of the data from the hydrophone array. However, efficiency should be considered at other sites which don't confer the same advantages. For example, tagging animals in areas where real Navy sources are about to operate may not be very lucrative in settings where tagged animals may move on before sound exposure starts. Surface observations will perhaps be more helpful in these situations.

PAM studies allow detection of vocalizing animals and can be useful to monitor marine mammal presence in areas of interest. They certainly add value to the marine mammal studies tool box.[13]These methods in deliberate combination with experimental methods can provide useful linkages from individual responses to population level estimates. The long-term nature of some data-sets may open up possibilities to fill knowledge gaps for a population consequences of acoustic disturbance (PCAD) model approach.

[5]PAM will need to be a component of any long-term study because it is less expensive and can provide data over longer time spans than even extended duration tags. PAM can provide information on distribution and abundance of species whose vocalizations can be discriminated. It can also track specific individuals in some cases. Obviously, it works only when the animals are vocalizing, and one of the initial responses to sonar has been to reduce or cease vocal activity. PAM can also be useful for studying dolphins and porpoises too small to be tagged. Algorithms to identify vocalizations to species for these cetaceans are still works in progress. Most of the pelagic delphinids cannot be automatically classified to species based on vocalizations. When combined with visual observations (such as the FLIP studies or focal-follow studies) group responses to sonar and other acoustic disturbance can be observed. Many of the results from visual observations will be more qualitative than quantitative, with focal-follow studies producing the most quantitative results.

Passive acoustics, particularly the passive acoustics associated with tracking beaked whales on naval testing ranges, has provided important data suggesting that the whales move off the range during anti-submarine warfare exercises but that when the exercises end, the whales return to the range. These studies have shown what is probably the most definitive evidence of potential long term impact of sonar activities by causing substantial temporal and spatial displacement of the animals. In such cases where PAM can be used for tracking animals we can be more confident that what is observed is actually a movement of the whales rather than a change in their vocal behavior.

PAM on the Navy ranges has been used to significantly advance our understanding of responses of range animals, particularly beaked whales. Given the depth of feeding and the directionality of beaked whale echolocation clicks, it is unlikely that the necessary data quality could be replicated by a portable PAM operating off the range, but there is still much data that can be collected through continuing PAM operations on the naval ranges.

Recommendations

[4 & 5]Much longer periods of baseline data should be obtained both on naval ranges and in relevant species habitats far from naval ranges. One would like to have baseline data at the same level of resolution as that provided by the D-tags in the CEE studies. However, for both the temporal and spatial extent of the desired baseline data, these tags will need to be supplemented by other approaches. Observational studies, particularly in conjunction with

PAM, can provide some baseline data but because indications are that the most important responses relate to changes in diving patterns and substantial horizontal displacement, [17]much of the needed baseline data will come from long-term Fastloc GPS satellite tags with the ability to transmit compressed dive profile data.

[4 & 5]Long term monitoring of individuals after exposure should be a priority to understand the time window that is affected by sound exposure. In birds, a brief playback of an intruding male can affect singing behavior over up to 24 hours (Amrhein & Erne 2006). This kind of information is important to have when studying the effects of noise, and will allow predicting longer-term and population level effects more clearly than the short term observations that are available now. Existing satellite tag work has started to look at this time line, but needs to be augmented with new tags that can measure sound exposure.[18]Therefore, a technological development that would be helpful in this study program is a longer term attachment of an acoustic tag.

The Fastloc GPS ARGOS satellite tags with either dive profile or acoustic sensors need long-term attachment to the animal. Parallel research endeavors will be needed to develop these tags with the appropriate sensors and to design [19 & 20]external attachment methods for use on smaller pelagic dolphins and implantable attachments for larger cetaceans.

[12]As discussed with reference to the CEE approach with real naval sonar, several studies have shown that both distance and received levels are important predictors of response. Whilst the best data will come from CEE in conjunction with Navy ships, it is likely that much of these data will need to be derived from observation of tagged animals in less controlled encounters.

[14]Ideally, PAM should use an array that allows tracking or some other way of assessing the number of animals vocalizing beyond just counting vocalizations. Early studies have shown that vocal activity may stop in response to noise exposure. This can either be caused by animals moving away or by ceasing to vocalize. Data from tagged animals will allow us to interpret such reactions. Once it is established what explanation accounts for vocalizations to stop, PAM may be used to look at reactions in animals that normally vocalize continuously. When large arrays are available in areas with resident populations, individual tracking may be possible as well as using individually distinctive parameters in vocalizations.

[6]Observational data can be collected from fixed observation points or during focal follows. The latter method has been underutilized and should be used to provide more information in the future. Observations from fixed points seem less useful for studying

reactions except for initial responses. All behavioral follows should focus on individuals since group sampling is highly problematic (Mann 1999).

One encouraging finding has been the prevalence of various species of marine mammals, including beaked whales most sensitive to stranding in the presence of sonar, on naval training ranges. [15]However, before this can build confidence in the long term coexistence of marine mammals and sonar, comparative demographics between populations resident on naval ranges and those in more pristine environments are needed to answer the question of whether the populations on naval ranges are a sink for neighboring populations. A recent paper (Whitehead and Gero 2015) suggests that the sperm whale population in the eastern Caribbean, although increasing in numbers and apparently healthy, is a sink due to probable human-caused mortality in this population. This documentation of a sink population gives much more substance to the hypothetical discussion and emphasizes the need for such a comparison of populations on and off naval training ranges.

Predator playback studies

Introduction

The leading hypothesis for behavioral responses to sonar sounds is that these stimuli evoke an anti-predator response, and that the probability of response is related to predation risk. If this hypothesis is supported by data then this could allow prediction of responsiveness in unstudied species. This has motivated research into how individuals respond to the sounds of predators and a comparison with the responses observed in response to sonar sounds. This work has, like the sonar CEE work, been conducted in an experimental context with playbacks of control stimuli and predator stimuli (namely killer whale calls). The work to date has been carried out on free-ranging individuals in the wild. Species studied include long-finned and short-finned pilot whales, humpback whales and sperm whales (e.g., Curé et al. 2012, 2013, 2015). This approach shares many of the same limitations as the CEEs using sonar, particularly the potential for research activities to confound the interpretation of the results.

Reviewer Comments

The results of most of the studies conducted so far have been consistent with the hypothesis that the responding animals perceive the sonar source as a potential predator. For most species the responses range from a cessation of vocalization to an interruption of foraging to leaving the area of the stimulus. Enhanced alertness, interruption of foraging and exiting from the area are normal responses to a predator. The response of pilot whales, while different from those of most species by approaching rather than leaving the area of the stimulus, does match their normal predator response which is to “mob” a killer whale predator. The most important outstanding questions relate to whether these normal predator responses will have individual and population-level consequences because of their frequency, intensity or cumulative effects.

The killer whale playbacks are believed to be a valuable approach, but what the consequences are to the animal remains an issue. It is a valuable approach if the aim is to find out what the short-term and long-term consequences are of the response in terms of energetic expenditure, stress physiological costs, lost foraging opportunities or social separations. However, it is not clear whether we gain any insight into the costs or detrimental impact of behavioral response if there is or is not a match to behavioral patterns under naturally risky circumstances. It may be, for example, that prey species that often shift rapidly in behavioral patterns, due to frequent potential approaches of predators, hardly suffer from changing their behavior once or twice more due to sonar exposure. Other species that are less vulnerable naturally may experience more detrimental impact due to the same or less dramatic behavioral responses.

Recommendations for future research

The external reviewers each contributed a number of high-level recommendations for future research that encompass all research approaches and are relevant to the overall research program looking at the effect of naval source on the behavior of marine mammals. Recommendations are based on the reviewers' views of the most important scientific questions required for Navy stewardship of oceanic marine mammals and responsiveness to regulatory requirements. These recommendations primarily result from listening to, and participating in, discussions during the workshop. The recommendations have been synthesized and summarized by the report authors and we do not repeat overlapping recommendations; rather, we have provided a list of the recommendations for future research along with rationale. We begin by outlining species priorities for future research and then move on to key topics that, with further research effort, could enhance our understanding of individual-level behavioral responses to naval sonar exposure and the potential for population-level consequences. These recommendations are listed in no particular order of priority. Rather we have included a summary table of all recommendations where a priority score has been assigned by the reviewers (Tables 1 and 2), and superscript numbers throughout the text link to the recommendations in Table 2.

Species priorities for future research.

Because beaked whales are the primary animals with documented individual, if not population, consequences through stranding, it is important to continue to place substantial effort in tagging and tracking these animals in spite of the difficulty in conducting such research (Table 1). Sample sizes for the most vulnerable species, such as the beaked whales, are generally too low. While these may be sufficient for a particular statistical method, they are too low to represent natural variation in an adequate way. The interplay between range and received level and response, including severity of response, should be investigated in beaked whales. Demographic profiles of naval training range populations and other populations can be investigated through visual observations of beaked whales.

While it is generally accepted that beaked whales are of particular concern, many other species remain untested. Therefore we also recommend research efforts on the following species and groups for the reasons given. ESA listed baleen whales and sperm whales are more amenable to tagging and tracking than beaked whales, and because of demonstrated short-term interruption of foraging they also have the potential to experience long-term impacts (Table 1). These species are appropriate for looking at body condition issues via DTAGS. Harbor porpoise have occasionally stranded and in general show the greatest sensitivity to acoustic disturbance (Table 1). A lot of good data has been obtained from captive animals but so far nothing from wild animals. [19]A Fastloc GPS tag with a dive profiler that could be attached to the dorsal fin of wild captured harbor porpoises should be a technological design goal. Such a tag would also be very useful in a similar attachment to a variety of oceanic delphinids that dominate take numbers because of their population size but for which we have virtually no information on either short-term or long-term impacts. Research on small delphinids should be a high priority given the number exposed each year (Table 1). Some delphinids observed passively appear to be more tolerant of acoustic disturbance than the whales studied to date although they do respond when received levels become high.

Baseline studies and long-term monitoring

In almost all studies, there has been significant variability in the response of the animals to very similar received levels. It should not be surprising that activities such as social behavior, travelling, foraging, and diurnal and annual cycles will influence the responsiveness of individuals. How important various activities are will vary over the course of a day or year, and the time spent in various activities will change over the course of a day or year. [4 & 5]Baseline studies are needed to begin to understand the non-stimulus based components of behavior. Data obtained so far indicate that seasonal and inter-annual oceanographic variability impacts behavior and habitat use more than the presence of sonar. Similarly, zebrafish studies in captivity and seabass studies in captivity (Neo et al. 2014, 2015) and in large outdoor enclosures (Neo et al. in preparation) show highly variable baseline patterns between and within groups (even in their simple and restricted captive environments). In general and independent of species, this causes problems in finding significant changes and elevates the required sample size.

The NRC report (NRC 2005) on the biological significance of ocean noise recommended comparing observed responses against the baseline data and determining where on the spectrum of baseline behaviors the observed response fell. As a conservative suggestion for determining biological significance, the report identified the 25th or 75th percentile of normal behavior. For example, it suggested there could be a biologically significant concern if the duration or length of the migration of an exposed animal was greater than the 75th percentile of normal migration time or length. For significant foraging impact, the report suggested the body condition of the animal would be below the 25th percentile of un-impacted animals at an equivalent time in the annual cycle. Whether or not the suggested percentiles of the NRC report are adopted, it does show the necessity of obtaining sufficient baseline data to make a biologically significant impact determination. A focus on baseline data is therefore advisable as baseline ecological studies of multiple species are key to our general understanding and the correct interpretation of both behavioral patterns and experimental results.

[4]In addition, long term monitoring of individuals after an exposure period should be a priority to improve our understanding of the time window affected by sound exposure. This kind of information is important when studying the effects of noise, and will allow predicting longer-term and population-level effects more clearly than the short term observations that are available now. [17-20]Existing satellite tag work has started to look at this time line, but needs to be augmented with new tags that can measure sound exposure and can provide dive profile data. The reviewers had differing opinions on which of these technological developments is higher priority, but did agree that the ideal scenario is for both to be developed in parallel.

Underlying mechanisms behind behavioral responses

Stranding events observed around some Navy exercises are clearly not adaptive. It is difficult to see how a mechanism that ensures fast spatial avoidance as an adaptive response to a predator or a threat in response to a signal as shown in some BRS studies would also account for animals swimming onto a beach as found in Navy exercises. There is a clear difference between those types of responses, which have been observed at similar received levels. This suggests that there might be two different mechanisms at work, one that explains behavioral responses such as avoidance, while a different one, such as the sensitization to startling stimuli as demonstrated by Goetz & Janik (2011), may explain stranding events. [21]An exposure to higher levels than used in BRS so far would be desirable, either in an experimental setting or by monitoring animals around an actual Navy exercise where high levels are more prevalent. It would allow us to see whether more extreme reactions develop out of low level avoidance, or whether there is a sudden shift in reactions at a particular threshold. Understanding what mechanisms are at work is the key to being able to predict reactions when operating around marine mammals and produce informed guidelines regarding when and how to use sonar around marine mammals to avoid lethal events.

BRS studies generally assumed that only one mechanism is responsible for behavioral reactions with extreme events occurring when the playback signal exceeds received levels above a certain threshold. An example is the assumption that the reaction represents predator avoidance that goes to an extreme if the threat appears much louder than previously experienced. This is focusing on the received level, but reactions may be caused or modified by other parameters. A startle response, for example, only occurs when a received level is supplemented by a short onset time. A reaction to a startle stimulus at low levels, (i.e., less than 90 dB above the hearing threshold) may be habituation, while the reaction when exceeding this level may turn into the opposite when animals are sensitizing to it. At low levels, received level might be responsible for the reaction, but at high levels it could be the combination of received level and onset time. While received level is clearly a very important aspect, [1]more parameters need to be tested for their potential to modulate responses. For example, in most cases, it appears that the playback of an actual predator, a killer whale, elicits a response at a lower level and a more sustained response than exposure to sonar. The sonar elicits a response at a somewhat lower response level than pseudorandom noise of similar timing and bandwidth, although the difference between response thresholds of sonar and pseudorandom noise are not nearly as great as those between killer whale vocalizations and either sonar or pseudorandom noise. [1]Therefore signal characteristics need to be further explored. Captive choice studies can be useful here in which animals are presented with different sounds in different parts or pools of their enclosure to observe where they choose to spend more time indicating a greater tolerance for the associated signal.

Appropriate response indicators

Captive dose-response studies have shown species differences in shape of the curve (symmetric or asymmetric), in age-related sensitivity, and in the occurrence of habituation. There has not been much overlap between the species studied in captivity and the species studied in the open ocean. In one case where the same species, bottlenose dolphin, was studied in captivity and in the wild, the captive studies would suggest that dolphins would show greater behavioral response to sonar than was shown by wild bottlenose dolphins on the PMRF range.

[22]Studies with killer whales have shown that sensation level is a more appropriate metric for measuring dose-response than received level. [21]This suggests that work should continue to obtain audiograms of as many potentially impacted species as possible to

generalize from received level to sensation level to probability of response. We still have no audiograms for any baleen whale.

[25]When severity of response was considered, there was not a clear dose-response relationship (Miller et al. 2012). Higher severity responses could occur at low dose and vice versa. Hence more data are needed here as well in order to begin drawing conclusions regarding mid-term and long-term impacts.

The relationship between source distance and received level

Most of the studies have shown that some animals start responding at received levels substantially below those requiring mitigation under current regulations. Some studies have shown that animals responding at low received levels to simulated sonar do not respond to significantly higher received levels from actual sonar operations where the naval ship source is much further away than the simulated sonar source. [11 & 12]The relationship between range and received level needs further investigation. The challenge of any experiment addressing this issue will be the sample size and distance control between animals and sound source.

Cumulative effects in noise exposure

[7 & 8]Cumulative effects may explain some of the more dramatic reactions to noise. These can take various forms. Repeated exposure has been tested in many BRS studies by exposing the animals over a period of time. What is lacking are studies with two sources, or repeated exposure of the same individuals over several days. Similarly, different source movements may lead to different reactions. Sometimes animals may get chased without operators being aware of this happening. This may lead to a more pronounced effect.

The role of experience

There is a critical requirement for any study that the sample population should represent the population for which we raised the question in the first place or for which we want to make a statement. This means that if we like to make a statement about all individuals of a particular species of beaked whales, some of which live in areas that are often exposed to sonar and some of which live in other areas that are less exposed, that investigations in just one particular area do not suffice. [15 & 16]Comparisons of animals that live in areas with frequent Navy exercises and those that live in quieter environments may help to understand the role of experience in noise exposure. Similarly, one could compare the reactions of young and old animals to the same stimuli or conduct longitudinal studies on the same individuals in resident populations. In elephants, experience has a major influence on how a group reacts to an acoustic challenge (McComb et al. 2001). A similar role of old animals in assessing threats is likely to be found in cetaceans.

The role of context

Context clearly modifies how animals respond to noise. This has been demonstrated already in case studies from existing BRS studies. It is possible that animals suffer long-term consequences when exposed to noise, but remain in an area because it is a lucrative foraging ground (e.g., Hastie et al 2015). Ultimately, these animals may still be compromised by the exposure to noise. Preliminary data from blue whales suggests that behavioral state is an even more important factor in determining response to a given dose than it is for odontocetes, where it is still a significant factor (Goldbogen et al. 2013). In some species, fleeing behavior can have a strong social component and so interactions among individuals may also explain why behavioral effects escalate in some situations and fade out quickly in others. [26]Tagging of multiple individuals simultaneously may reveal such interactions and also allow the study of individual differences to the very same exposure situation. Using the real Navy sources instead of scaled playbacks adds realism to these types of experiments, but control exposures to other sounds such as white noise or even other man-made sounds, such as pile driving or airguns may allow comparisons in the qualitative and quantitative nature of behavioral responses and put sonar into context. Studying the contexts in which animals react or do not react is crucial to understanding the impacts of sonar. [22 & 24]Across all species the data set linking behavioral state and response needs to be expanded.

Long term effects of noise exposure

Behavioral responses have been demonstrated in all species tested, albeit at variable received levels. Cutting-edge statistics have been developed and refined to demonstrate a statistically significant response in a number of cases. The changes in behavior observed clearly have the potential to have long-term impact as a disruption of foraging is a common response. Displacement is another response that takes energy and may move animals to areas where food resources are diminished. Against this potential however are some broad observations of populations apparently surviving well through decades of naval exercises on the ranges that overlap the home ranges of these populations. Setting aside the stranding response, are there population consequences of sonar exposure? [4 & 5]A first step in answering this question is to obtain data over a longer period of time, over greater spatial extent and to replicate across different regions and populations. Study replication is needed at the population level to really get evidence of an impact. PAM could be a valuable component of any long-term study because it is less expensive and can provide data over longer time spans than even extended duration tags.

Repeated exposure may lead to increased stress and ultimately to fitness consequences. [28]New approaches need to be explored to investigate such effects and to assess when sound exposure has the potential to become a problem at the population level. Studies addressing these questions would look at the effects of noise on stress levels, and how stress modifies behavioral responses. Measurable changes in stress hormones in faeces have been hypothesized to be associated with corresponding changes in ship traffic in the Bay of Fundy (Rolland et al. 2012). So far little work has been done relative to stress hormones and exposure to sonar but looking at stress hormones could provide information on long-term effects. Fecal stress hormones have been used to track short-term changes in stress. Biopsies have been faulted because of the long-term integration of stress signals but such integration is what is needed for a comparison between animals exposed to sonar on a regular basis and animals living in more pristine conditions. Both fecal stress hormones and stress hormone signatures in skin and blubber should be assessed for possible differences related to exposure to sonar.

Energetics and body condition assessment are also key components of long-term impact assessment. Energy has proven to be a useful currency in development of the PCAD/PCoD model. [29]Body condition studies may allow quantifying long term changes when comparing animals at the start and the end of an exposure period in a chosen location, or in observational studies when comparing body condition in exposure zones such as on the Navy ranges to those of animals that live in areas with little sound exposure. A change in body condition is clearly not a behavioral response itself, but could result from a behavioral response and could also lead to behavioral modifications due to the resulting changes in energetic requirements. It is possible that aspects of body condition can be evaluated through swimming and diving/drifting/sinking/rising data obtained with the accelerometers on the DTAG. By investigating these relationships, we may be able to predict long term effects by observing initial responses to noise exposure.

[15]Demographic studies that allow comparison of populations resident on naval ranges and those in more pristine environments can help identify potential long-term effects of exposure on population dynamics, and, in naval range areas where animal prevalence remains

high despite exposure, they can help to answer the question of whether these populations are sinks for neighboring populations.

The study of long-term impacts requires some technological developments. [19 & 20]Tag attachment duration must be improved. The attachment time needs to progress from hours to at least a week and preferably longer. Such an attachment will require either a limpet attachment or some type of fully implanted tag. Because most of the BRS work is being conducted at low latitudes where ARGOS tracking is limited, this tag will need to incorporate Fastloc GPS. [17 & 18]After location, the next most important sensors are a depth-time sensor and a hydrophone. The depth-time sensor needs to have appropriate processing capabilities to send summarized dive profile information during satellite data transmission. The hydrophone would require associated processing so animal vocalizations and ambient sounds can be summarized and transmitted. Such a tag does not necessarily need to record full bandwidth audio continuously. A tag that processes acoustic input and provides a cumulative or count measure of sound exposure would be helpful to look at longer term and cumulative effects. It is understood that a medium-duration Fastloc GPS location only satellite tag is under development for possible deployment in early 2016.

Statistical review

Whilst the reviewers felt able to evaluate the crude assumptions of the statistical analyses carried out by the various BRS projects, they did not feel qualified to fully evaluate the range of statistical approaches developed for application to BRS data. While the empirical data are collected by a range of research groups and PIs, the statistical treatment of data appears to be combined in only one group. [30]A review of statistical methods by an independent reviewer would therefore perhaps be advisable.

Prey Response

[27]A PCAD model approach not only requires insight into natural and modified patterns of swimming and diving, but also on whether foraging opportunities have changed due to these altered patterns and whether foraging efficiency is affected in any way. There are few papers in which we have detailed knowledge on prey abundance and quality in the area that was left behind as well as in the alternate area. Furthermore, prey may respond to sonar and alter water column depth or swimming patterns and the presence of sound may also alter the foraging efficiency. The consequences of sonar exposure events may have immediate, as well as longer lasting, spatial effects on future travelling routes depending on the degree of

anxiety caused by the event itself and the altered experience in terms of feeding rates and quality.

From what we have already learned indirectly about prey response, it appears thus far to be less sensitive than the marine mammal response. There are several observations of whales experiencing higher received levels before responding when they are feeding than when they are engaged in other activities with the implication that the whales remained because the prey continued to be available. Any sonar effects on prey availability need to be incorporated in the cumulative effects of sonar on marine mammals, but the response of the prey to sonar is likely to be a second order effect at the current level of analysis. [27]It was recommended to explore possibilities for studies in which the ecological consequences of behavioral effects are measured and concurrent investigation of dominant prey, predator or competitor species may provide a logical approach.

Ramp-up

[31]Calculations based on studies to date indicate that under certain narrowly defined circumstances ramp-up could reduce the exposure levels of animals, provided they responded to the ramp-up appropriately. However, in the case of naval sonar the speed of naval ships is such that ramp-up will have little effect on SEL.

Regulation

While not specifically a research topic, the findings to date with the CEE have demonstrated that NMFS and the Navy need to look at regulatory and possibly legislative changes so that take becomes a more scientifically justifiable concept. Having demonstrated statistically significant behavioral changes occur in some species and for some signals at the level of audibility, the current application of take will result in huge numbers of animals for which monitoring and mitigation would be impossible.

The Navy could make use of the revised Level B harassment definition provided in 16 U.S.C. 1362 (18)(B)(ii) where Level B harassment for military operations is defined as responses that occur "to a point where such behavioral patterns are abandoned or significantly altered." It is of course subject to interpretation as how long a given behavior (e.g., foraging) is interrupted before meeting the definition of being abandoned. Similarly, significantly altered could be interpreted in a statistically significant sense or in a biologically significant sense.

The experiments have shown that some form of a dose-response curve, rather than a step function, best describes the impact of sonar on all species studied to date. Thus a scientifically based take would be a number with a confidence interval. The continuation of presenting a single number for the take of any species is not the best available science.

Concluding comments

In conclusion, we stress the importance of understanding the general ecology of marine mammals and especially the role of sounds in their life. Interpretation of behavioral responses requires a thorough understanding of physiological and ecological consequences, which relies heavily on knowledge about what the animals typically do and would have done otherwise in the absence of sonar and in the absence of human researchers. [4 & 5] Baseline ecological studies of multiple species are key to our general understanding and the correct interpretation of both behavioral patterns and experimental results.

The problem of understanding behavioral patterns and especially event-related deviations is not trivial in natural environments. One can perhaps learn from experience with other taxa, such as with captive fish that can be observed continuously before, during and after sound exposure and for which there are also data on how they respond to visual threats from above, the visual presence of an actual predator fish, or alarm pheromone in the water that is indicative of a nearby and active predator.

All marine mammal studies, captive and free-ranging, have shown that responses of animals are highly variable. The difficulty in drawing conclusions from data that are highly variable based on largely unknown intrinsic and extrinsic factors has required the development of new analytical techniques and the application of the latest statistical techniques to marine mammal behavioral responses. These techniques such as Mahalanobis distance, hidden state models, recurrent event survival analysis, and Bayesian hierarchical models can have broad applicability beyond the current sonar response studies.

Given the variability in responses there are questions regarding the applicability of dose-response curves created with one species under one set of conditions to other species and situations. One of the contributions of the program to date has been the clear evidence of the fallacy of attempting to provide one number for the potential takes in any operation. Operators and regulators need to work with legislative bodies to embed the inherent uncertainty of any take estimates in the regulatory framework.

We still do not have sufficient data to be able to extrapolate from the current set of experimental subjects to broader conclusions about which species may be more sensitive to, or more tolerant of, sonar. Based on the small number of species looked at so far, the classification of cetaceans based on functional hearing groups according to the general hearing ranges does not appear to be appropriate for assessing sonar impacts. In the class of mid-frequency cetaceans, bottlenose dolphins (at least on Navy ranges) and pilot whales appear to be relatively tolerant of sonar, with pilot whales attracted to playback sources, whereas sperm and killer whales show increasing levels of response and beaked whales are the most responsive. When more data are available from a broader range of species, new functional groupings may become more apparent.

Finally, translating the results from any of the studies conducted to free-ranging animals exposed under real exercise conditions should be carried out with caution. All studies are limited in how well their study animals represent the animals and conditions of the target populations for risk assessment. In particular many of the studies so far are limited by lack of replication. Statements at the population level are only solid in terms of causation when replicated at the population level. Comparing one population with another, or the same population in two different time periods, reflects a sample size of one. As sample sizes are typically limited and restricted to particular areas, seasons and test conditions, this is a concern for all sonar-impact studies.

Acknowledgements

The workshop and report were jointly sponsored by the Office of Naval Research, Navy Living Marine Resources, and the National Oceanic and Atmospheric Administration - National Marine Fisheries Service. The workshop venue was kindly provided by Hopkins Marine Laboratory.

References

Amrhein, V. & Erne, N. 2006. Dawn singing reflects past territorial challenges in the winter wren. Animal Behavior 71**,** 1075-1080.

Antunes, R., Kvadsheim, P.H., Lam, F.P.A., Tyack, P.L., Thomas, L., Wensveen, P.J., & Miller, P. J. O. 2014. High response thresholds for avoidance of sonar by free-ranging long-finned pilot whales (*Globicephalamelas*). Marine Pollution Bulletin (DOI: 10.1016/j.marpolbul.2014.03.056)

Curé, C., Doksaeter-Sivle, L., Visser, F., Wensveen, P.J., Isojunno, S., Harris, C.M., Kvadsheim, P., Lam, F.P.A. & Miller, P.J.O. 2015. Predator sound playbacks reveal strong avoidance responses in a fight strategist baleen whale. Marine Ecology Progress Series 526, 267-282. http://dx.doi.org/10.3354/meps11231.

Curé, C., Antunes, R., Alves, A.C., Visser, F., Kvadsheim, P.H. & Miller, P.J.O. 2013. Responses of male sperm whales (*Physetermacrocephalus*) to killer whale sounds: implications for anti-predator strategies. Nature Scientific Reports 3 (1579): 1-7.

Curé, C., Antunes, R., Samarra, F., Alves, AC., Visser, F., Kvasheim, P.H. & Miller, P.J.O. 2012. Pilot whales attracted to killer whales sounds: acoustically-mediated interspecific interactions in cetaceans. Plos One 7(12): e52201.

D’Amico, A., Gisiner, R.C., Ketten, D.R., Hammock, J.A., Johnson, C., Tyack, P.L. & Mead, J. 2009. Beaked whale strandings and naval exercises. Aquatic Mammals 35:452-472.

DeRuiter, S. L. 2010. Marine Animal Acoustics. In *An Introduction to Underwater Acoustics* (ed X. Lurton), pp. 425–474. Chichester, UK: Praxis Publishing Limited.

DeRuiter, S.L., Southall, B.L., Calambokidis, J., Zimmer, W.M.X., Sadykova, D., Falcone, E.A., Friedlaender, A.S., Joseph, J.E., Moretti, D., Schorr, G.S., Thomas, L. & Tyack, P.L. 2013. First direct measurements of behavioral responses by Cuvier’s beaked whales to mid-frequency active (MFA) sonar. Biology Letters 9: 20130223.

Goldbogen, J.A., Southall, B.L., DeRuiter, S.L., Calambokidis, J., Friedlaender, A.S., Hazen, E.L., Falcone, E.A., Schorr, G.S., Douglas, A., Moretti, D.J., Kyburg, C., McKenna, M.F. & Tyack, P.L. 2013. Blue whales respond to simulated mid-frequency military sonar. Proceedings of the Royal Society B. 280(1765), 20130657. doi:10.1098/rspb.2013.0657

Götz, T. & Janik, V.M. 2011. Repeated elicitation of the acoustic startle reflex leads to sensitisation in subsequent avoidance behavior and induces fear conditioning. BMC Neuroscience 12**,** 30.

Harris, CM, Sadykova, D, De Ruiter, SL, Tyack, PL, Miller, P, Kvadsheim, P, Lam, F.P. & Thomas, L. 2015. Dose response severity functions for acoustic disturbance in cetaceans using recurrent event survival analysis. Ecosphere (in press).

Hastie, G.D., Russell, D.J.F., McConnell, B., Moss, S., Thompson, D. & Janik, V.M. 2015. Sound exposure in harbour seals during the installation of an offshore wind farm: predictions of auditory damage. Journal of Applied Ecology 52, 631-640.

Houser, D., Yeates, L., Crocker, D., Martin, S.W. & Finneran, J.J. 2012.Controlled exposure study of dolphins and sea lions to mid-frequency sonar-like signals. Advances in Experimental Medicine and Biology 730, DOI 10.1007/978-1-4419-7311-5_59, A.N. Popper and A. Hawkins (eds.), The Effects of Noise on Aquatic Life.

Houser, D.S., Martin, S.W.& Finneran, J.J. 2013. Exposure amplitude and repetition affect bottlenose dolphin behavioral responses to simulated mid-frequency sonar signals. Journal of Experimental Marine Biology and Ecology, 443, 123-133.

Houser, D.S., Martin, S.W. & Finneran, J.J. 2013.Behavioral responses of California sea lions to mid-frequency (3250–3450 Hz) sonar signals, Marine Environmental Research, 92, 268-278.

Kastelein, R. A., Steen, N., de Jong, C., Wensveen, P. J., & Verboom, W. C. 2011. Effect of broadband-noise masking on the behavioral response of a harbor porpoise (*Phocoena phocoena*) to 1-s duration 6-7 kHz sonar up-sweeps. Journal of the Acoustical Society of America, 129, 2307–2315. doi:10.1121/1.3559679

Kastelein, R.A., Gransier, R., van den Hoogen, M. & Hoek, L. 2013. Brief Behavioral Response Threshold Levels of a Harbor Porpoise (Phocoenaphocoena) to Five Helicopter Dipping Sonar Signals (1.33 to 1.43 kHz). Aquatic Mammals, 39(2), 162-173.

Mann, J. 1999. Behavioral sampling methods for cetaceans: a review and critique. Marine Mammal Science.15, 102-122.

McComb, K., Moss, C., Durant, S.M., Baker, L. & Sayialel, S. 2001. Matriarchs as repositories of social knowledge in African elephants. Science 292**,** 491-494.

Miller, P.J.O., Johnson, M.P., Madsen, P.T., Biassoni, N., Quero, M., & Tyack, P.L. 2009. Using at-sea experiments to study the effects of airguns on the foraging behavior of sperm whales in the Gulf of Mexico. Deep-Sea Research I 56, 1168-1181.

Miller. P.J.O., Kvadsheim, P.H., Lam., F.P.A., Wensveen, P.J., Antunes, R., Alves, A.C., Visser, F., Kleivane, L., Tyack, P.L. & Doksæter-Sivle. L. 2012. The severity of behavioral changes observed during experimental exposures of killer (*Orcinus orca*), long-finned pilot (*Globicephala melas*), and sperm (*Physeter macrocephalus*) whales to naval sonar. Aquatic Mammals 38(4): 362-401.

Miller, P.J.O., R.N. Antunes, P.J. Wensveen, F.I.P. Samarra, A. Catarina Alves, P.H. Kvadsheim, L. Kleivana, F.-P.A. Lam, M.A. Ainsle, P.L. Tyack & L. Thomas. 2014. Dose-response relationships for the onset of avoidance of sonar by free-ranging killer whales. Journal of the Acoustical Society of America.135, 975. http://dx.doi.org/10.1121/1.4861346

Miller P.J.O., Kvadsheim, P.H., Lam, F.-P.A., Tyack, P.L., Curé, C., DeRuiter, S.L., Kleivane, L., Sivle, L.D., van Ijsselmuide, S.P., Visser, F., Wensveen, P.J., & Hooker, S.K. 2015. Northern bottlenose whales are highly sensitive to noise disturbance. Royal Society Open Science.2: 140484.

National Research Council. 2005. Marine Mammal Populations and Ocean Noise: Determining When Noise Causes Biologically Significant Effects. National Academies Press, Washington, DC

Neo, Y.Y., Parie, L., Bakker, F., Snelderwaard, P., Tudorache, C., Schaaf, M. & Slabbekoorn, H. 2015. Behavioral changes in response to sound exposure and no spatial avoidance of noisy conditions in captive zebrafish. Frontiers in Behavioral Neuroscience 9:1–11.

Neo, Y.Y., Seitz, J., Kastelein, R., Winter, H.V., ten Cate C. & Slabbekoorn, H. 2014. Temporal structure of sound affects behavioural recovery from noise impact in European seabass. Biological Conservation 178:65–73.

Richardson, W. J., Greene Jr., C. R., Malme, C. I. & Thompson, D. 1995 *Marine Mammals and Noise*. San Diego, CA: Academic Press.

Rosalind M. Rolland, Susan E. Parks, Kathleen E. Hunt, Manuel Castellote, Peter J. Corkeron, Douglas P. Nowacek, Samuel K. Wasser, Scott D. Kraus

Rolland, R.M., Parks, S.E., Hunt, K.E., Castellote, M., Corkeron, P.J., Nowacek, D.P., Wasser, S.K. & Kraus, S.D. 2012. Evidence that ship noise increases stress in right whales. Proceedings of the Royal Society B. 279:2363–2368

Schorr, G.S., Falcone, E.A., Moretti, D.J. & Andrews, R.D. 2014.First long-term behavioral records from Cuvier's beaked whales (*Ziphius cavirostris*) reveal record-breaking dives. PLoS ONE 9(3):e92633.doi: 10.1371/journal.pone.0092633

Tyack, P.L., Zimmer, W.M.X., Moretti, D., Southall, B.L., Claridge, D.E., Durban, J.W., Clark, C.W., D'Amico, A., Di Marzio, N, Jarvis, S., McCarthy, E., Morrissey, R., Ward, J. & Boyd, I.L. 2011. Beaked whales respond to simulated and actual navy sonar. 2011. PLoS One 6(3): e17009

Whitehead, H. & Gero, S. 2015. Conflicting rates of increase in the sperm whale population of the eastern Caribbean: positive observed rates do not reflect a healthy population. Endangered Species Research 27: 207–218.

Table 1: Species priorities for future research.

ID	Species/species group	Priority score
A	Beaked whales	1
B	ESA listed baleen whales	2
C	Sperm whales	2
D	Harbour porpoise	2
E	Small delphinids	2

Table 2: Recommendations for future research effort (order of recommendations approximately relates to order of appearance in report text). All recommendations are linked to the text by the ID number, which can be found next to the relevant sections in superscript. The priority score is made up of three independent scores provided by the reviewers. Each reviewer gave each recommendation a score of 1, 2 or 3 with a score of 1 indicating the highest priority level. The combined score is ordered with the highest priority score first.

Note: (1) Even those topics that scored a 3 are considered a priority. Topics that are not a priority are not listed. (2) In the assignment of priority, reviewers largely focussed on scientific reward and largely ignored feasibility. We indicate cases where priority has been adversely affected by the belief that the recommendation will only be feasible in the very long term. (3) All reviewers tried to follow an advised distribution of scores (20% rank 1, 70% rank 2, 10% rank 3).

ID	Recommendation	Approach	Priority score	Note
1	Captive choice studies to investigate: -parameters (other than RL) that may modulate response, including signal characteristics -response mechanisms -tolerance, preferences and aversion to acoustic stimuli	Captive	1-2-2	
2	Captive studies to investigate effect of noise in undisturbed animals which are carrying out normal activities	Captive	1-2-3	Not specific to sonar, but essential to not misinterpret lack of deterrence as lack of impact; also critical for PCAD-models
3	Test new technologies, such as physiology tags, on captive and free-ranging animals[1]	Captive	2-2-2	"Tagnological" advancements will be critical for any major advancements in understanding impact
4	Increase collection of baseline data (improve both spatial and temporal extent of data)	Free-ranging CEEs and observational with sat tags	1-1-1	

5	Increase collection of baseline data (improve both spatial and temporal extent of data)	Observational with PAM	2-2-2	
6	Increase use of individual focal follow observations	Free-ranging CEE and observational	2-2-2	Could be used to reduce reliance on tags and increase sample size associated with CEEs for some species. This can yield better understanding, especially if combined with advanced tags
7	Investigation of cumulative effects using two sources or repeated exposures over multiple days	Free-ranging CEEs	2-2-3	Important but exceedingly difficult. Concern that this will not become well-replicated
8	Investigation of source vessel orientation and movement	Free-ranging CEEs	2-2-2	
9	Investigate species differences in individual variability	Free-ranging CEEs	2-2-2	
10	Conducting CEEs with real Navy vessels and sources to generate dose-response functions for real sonar sources	Free-ranging CEEs with real sources	1-1-2	Requires adequate replication and appropriate sampling design
11	Investigate the relationship between source-whale distance and received level	Free-ranging CEEs with real sources	1-1-2	Logistically difficult, although not as difficult as two sources or repeated multiple days
12	Investigate the relationship between source-whale distance and received level	Observational with sat tags	2-2-2	
13	Integration of research approaches (e.g. PAM with CEEs)	All	2-2-3	
14	Improving interpretation of PAM data to increase potential for observing responses (i.e. move away or cease vocalisation)	PAM with tagging	2-3-3	The problem is that results cannot easily be applied to other locations and species. PAM is of limited use for looking at effects.
15	Comparison of population demographics in populations resident on naval ranges and those in areas far from naval ranges.	Observational (photo-id)	1-2-2	Required because of the political necessity of saying something about whether the range is a sink for other populations. Replication at popn

				level is required to prevent over-interpretation of a comparison of two populations
16	Conduct experiments in regions where animals are, and are not, likely to have previous experience of sonar exposure to understand role of experience.	Free-ranging CEE	1-2-2	Preferentially with long-term tag data on exposure history (for sonar and any other anthropogenic noise) for experimentally exposed whales
17	Development of a medium-long term Fastloc GPS tag with dive profiler	Technology	2-2-2	
18	Development of a medium-long term Fastloc GPS tag that can measure sound exposure	Technology	1-1-2	
19	Development of attachment mechanism to fit GPS tag (described in 15 and 16) to porpoises and small delphinids	Technology	1-1-2	
20	Development of long-term tag attachment mechanism for large species that cannot be captured for attachment.	Technology	1-1-2	
21	Conduct or observe exposures with higher received levels to better understand mechanisms underlying response	Observational near ranges, or free-ranging CEE	1-2-2	
22	Improve understanding of possible response metrics such as sensation level, range, behavioural state	All	2-2-2	
23	Obtain audiogram data for more species to allow investigation of sensation level as a response metric for more species		2-3-3	
24	Improve understanding of link between behaviour state and response	Observational and free-ranging CEE	1-2-2	
25	Improve understanding of link between dose and severity of response to better understand consequences	Observational and free-ranging CEE	1-1-2	With an emphasis on understanding consequences
26	Improve our understanding of social context by tagging multiple individuals simultaneously	Observational and free-ranging CEE	2-2-2	
27	Improve our understanding of the role of prey availability by collecting prey data alongside tagging	Observational and free-ranging CEE	1-2-2	This is critical for any interpretation with PCAD-models

	efforts, and looking at sonar effects on prey		
28	Improve understanding of long-term consequences through studying stress hormones of animals in different environments		2-2-2
29	Body condition studies in relation to exposure, e.g. using metrics from Dtags		2-2-2
30	Statistical review of the statistical methods developed for application to BRS		2-3-3
31	Investigate the effect of ramp-up protocols	Free-ranging CEE	2-3-3

[1]This recommendation would have been scored higher; however the reviewers felt that it will not be achievable in the short-term and so should be considered a longer-term goal.

Appendix 1 - Pre-workshop Questionnaire

Behavioral Response Research Evaluation Workshop (BRREW) - questionnaire.

The aim of this questionnaire is to elicit an assessment of the current state of knowledge in key areas, the research gaps, and suggested pathways to fill knowledge gaps. We anticipate it should take no more than two hours of your time. Please feel free to canvass opinions from others in your research group to help you complete this.
The results will be synthesized and circulated to all participants, including external reviewers, prior to the workshop. Individual responses will be anonymized.
The structure of the workshop will revolve around the main topics outlined below, and combined responses will be used to structure discussions.
If you do not feel that you can comment on a particular topic then please just state this.

Research topic 1: General overview questions

What, in your opinion, are the **key questions** for the US Navy relating to the behavioral responses of marine mammals to Naval sonar?

Please summarize your overall assessment of the current state of knowledge of marine mammal behavioral responses to Naval sonar:

Please provide details of any technological developments that you would like to see to help further this research area:

What additional studies on animals are required, in your opinion, to answer any remaining key questions?

Research topic 2: Responses to *simulated sources* of Navy sonar – Controlled Exposure Experiment (CEE) studies on *captive animals*

What are the **key research needs** that have been addressed using this approach?

What are the **limitations** associated with the research that has been done to date using this approach?

Please identify any **current research gaps** that could be filled using captive studies:

Please summarize your assessment of the **main limitations in meeting future research needs** given the current state of technology, experimental design and analytical tools in this field:

Research topic 3: Responses to *simulated* Navy sources *or* real but scaled Navy sources on research vessels – CEE studies on *free-ranging animals*

What are the **key research needs** that have been addressed using this approach?

What are the **limitations** associated with the research that has been done to date using this approach?

Please identify any **current research gaps** that could be filled using these kinds of studies:

Please summarize your assessment of the **main limitations in meeting future research needs** given the current state of technology, experimental design and analytical tools in this field:

Research topic 4: Responses to *real* Navy sources deployed by Navy vessels – *free-ranging* CEE studies

What are the **key research needs** that have been addressed using this approach?

What are the **limitations** associated with the research that has been done to date using this approach?

Please identify any **current research gaps** that could be filled using these studies:

Please summarize your assessment of the **main limitations in meeting future research needs** given the current state of technology, experimental design and analytical tools in this field:

Research topic 5: Responses to *real* Navy sources and Navy vessels – *observational* (e.g., opportunistic or incidental) studies

(Notes. 1. The key difference between this and topic 4 is that the researcher does not control the sonar operations, so it is not a *Controlled* Exposure Experiment. 2.We include here both short-term (single exposure) and long-term (multi-day) studies – please clarify which you are referring to in your responses.)

What are the **key research needs** that have been addressed using this approach?

What are the **limitations** associated with the research that has been done to date using this approach?

Please identify any **current research gaps** that could be filled using observational studies:

Please summarize your assessment of the **main limitations in meeting future research needs** given the current state of technology, experimental design and analytical tools in this field:

Research Topic 6: Interpretation of responses

Are predator playback studies useful in aiding our interpretation of sonar exposure experiments?

Do we have sufficient baseline data to understand whether behaviors observed during sonar exposure are truly “unusual”?

Can you suggest other studies that may help with interpretation of the data collected during sonar exposures?

Appendix 2 - Workshop agenda

BRREW project
Hopkins Marine Station, Monterey
21-22 April 2015

21 April

0900 Welcome and introduction to the workshop by the sponsors.

Research Topic 1 – What are the key questions for the US Navy?
0920 An overview of questionnaire responses, followed by discussion
1015 COFFEE

Research Topic 2: Responses to simulated sources of Navy sonar – Controlled Exposure Experiment (CEE) studies on captive animals
1045 An overview of questionnaire responses relating to captive CEE studies, followed by discussion
1215 LUNCH

Research topic 3: Responses to simulated Navy sources or real but scaled Navy sources on research vessels – CEE studies on free-ranging animals
1315 An overview of questionnaire responses relating to free-ranging CEE studies using simulated or scaled sources, followed by discussion
1445 COFFEE

Research topic 4: Responses to real Navy sources deployed by Navy vessels – free-ranging CEE studies
1500 An overview of questionnaire responses relating to free-ranging CEE studies using real Navy sources and vessels, followed by discussion

General discussion
1630 General discussion and Q&A session relating to any of the topics discussed during the course of the day
1715 Finish
1930 DINNER (DETAILS TBC)

22 April

Research topic 5: Responses to real Navy sources and Navy vessels – observational (e.g., opportunistic or incidental) studies
0830 An overview of questionnaire responses relating to observational studies using tags, followed by discussion
0930 An overview of questionnaire responses relating to observational studies using PAM, followed by discussion
1030 COFFEE

Research Topic 6: Interpretation of responses

1045 An overview of questionnaire responses relating to predator playback studies, followed by discussion

1115 An overview of questionnaire responses relating to baseline data, followed by discussion

1145 An overview of questionnaire responses relating to suggestions for other studies to aid with interpretation, followed by discussion

1215 LUNCH

Discussion session

1315 General discussion session, where we will present some responses from other general questions from questionnaire

1430 COFFEE

1445 Discuss and synthesize recommendations from each research topic

1700 Finish

Appendix 3 - Submitted Participant Biographies

Erin Falcone
Erin Falcone is a research biologist with Cascadia Research Collective- a small non-profit organization based in Olympia, Washington. Prior to joining Cascadia, Erin's research focused on reproductive strategies of humpback whales, operating as both a lead field technician and photo-identification data manager. This transitioned into her initial role at Cascadia in both data collection and as data manager for the SPLASH North Pacific humpback whale comprehensive assessment, the largest study of its kind. In 2006 she became a principle investigator for visual surveys associated with the Marine Mammal Monitoring at Ranges (M3R) acoustic monitoring program at the Southern California Offshore Range (SCORE). As part of this work, she initiated the first regional photo-identification studies of fin whales and Cuvier's beaked whales. This project has expanded to include satellite telemetry to collect extended movement and diving records from individuals in these strategic populations at SCORE, and most recently is combining these data products with records of sonar use to study both short-medium term behavioral changes, and potential longer-term demographic impacts, associated with real training exercises in the region. She also works as the visual survey data manager and a field technician deploying tags and collecting behavioral observations for the SOCAL BRS project.

Catriona Harris
Catriona Harris is a quantitative ecologist with experience working within the disciplines of marine mammal ecology, invasive species ecology and epidemiology. Catriona is a senior research fellow in the Centre for Research into Ecological and Environmental Modelling (CREEM) at the University of St Andrews, and divides her time between project managing high-profile commercial/research contracts and conducting her own research. Over the last 10 years she has been involved in a number of large projects relating to the impact of anthropogenic activities on marine mammals and the development of statistical methods for marine mammal detection and density estimation. Currently she is a Principal Investigator on an international collaborative project developing statistical methods to analyse data on the behavioral responses of cetaceans to acoustic disturbance.

John Harwood
John Harwood is a professor of biology at the University of St. Andrews. He is a former director of the Sea Mammal Research Unit, which advises the U.K. and Scottish governments on the conservation of seals and whales, and of the Centre for Research into Ecological and Environmental Modelling. His main research interest is in developing methods for assessing and mitigating the effects of anthropogenic disturbance on marine ecosystems. Additional research involves exploring the effects of individual variation and spatial structure on the population dynamics, genetics and epidemiology of vertebrates, particularly marine mammals. He is currently co-chair of ONR’s Population Consequences of Disturbance Working Group and a member of the National Research Council's Committee on Assessment of the Cumulative Effects of Anthropogenic Stressors on Marine Mammals.

John Hildebrand
Biography not provided.

Dorian Houser

Dorian Houser is the Director of Conservation and Biological Research at the National Marine Mammal Foundation. Dorian's research covers multiple aspects of marine mammal physiology, behavior and bioacoustics and he has published more than 90 peer-reviewed research articles and book chapters covering these topics. He was co-recipient of the Strategic Environmental Research and Development Program (SERDP) "Project of the Year" award in 2000 and received the R. Bruce Lindsay award from the Acoustical Society of America in 2007. He currently serves as the Vice-Chair of the Accredited Standards Committee (ASC) S3/SC 1, Animal Bioacoustics and is the incoming chair in 2015. He is also a member of ASC S3/WG2 Bioacoustics standards working group. Dorian has had significant involvement addressing marine mammal issues of concern to the Navy and has been involved in the development of numerous Navy environmental impact statements and environmental assessments involving marine mammals.

PetterKvadsheim
PetterKvadsheim is program manager of the Marine Environment research program at the Norwegian Defence Research Establishment (FFI). He has been working on research on effects of naval sonar on fish and marine mammals since 2003. Kvadsheim's academic background is physiological research on marine mammals from University of Tromsø, focusing on diving and thermoregulatory physiology. The past 10 years he has turned in focus more into behavioral biology. He is currently the chief scientist of the Sea Mammals and Sonar Safety (3S) project, doing field based behavioral response studies to six different species of cetaceans. Kvadsheim is also involved in the implementation of research on effects of sonar on marine life in operational use within the Norwegian Navy. He is also member of the Norwegian marine mammals commission with effects of anthropogenic noise as his speciality.

Frans-Peter Lam
Frans-Peter Lam holds a PhD in Physical Oceanography (Netherlands Institute for Sea Research NIOZ and Utrecht University). He is affiliated at the Acoustics & Sonar Research group of the Netherlands Institute for Applied Scientific Research (TNO) since 1998, where initially he contributed to the development and signal processing of Low Frequency Active Sonar (LFAS) systems. As a senior scientist he is currently leading the research program of effects of sonar on marine mammals. He is a board-member of the 3S consortium, one of the research groups that studies behavioral responses of marine mammals exposed to sonar. His main research interests are studying the effects of sound on marine mammals, acoustic detection and tracking of marine mammals and military oceanography (ocean forecasting), and has been visiting scientist at various institutes, such as Harvard, MIT, WHOI, DAMPT (Cambridge University) and NURC/CMRE. During the last 24 years, he has participated in over 20 sea trials (both on navy and research vessels), including regular sea trials in Norway since 2006 for the 3S controlled sonar exposure experiments. He is advising the Royal Netherlands Navy on potential impact of active sonar and was organizer of the "Effects of Sound in the Ocean on Marine Mammals (ESOMM)" international meetings in Amsterdam in 2011 and 2014.

Patrick Miller
Patrick Miller is a reader in the School of Biology at the University of St Andrews, and a member of the Sea Mammal Research Unit. Miller's research focuses on animal communication, behavioral ecology, kinematics, and body condition of cetaceans. Miller was awarded the 2013 Kobe Prize in Marine Science, for fundamental contributions to our understanding of the sperm whale. Miller has been involved in a number of studies of the

effects of noise on cetaceans, including: the LFA Scientific Resarch Program, the SWSS study of seismics and sperm whales in the Gulf of Mexico, and the ongoing 3S research collaboration studying the behavioral effects of sonar on cetaceans.

David Moretti
Biography not provided

Andrew Read
Andy is the Stephen Toth Professor of Marine Biology at the Duke University Marine Laboratory, in Beaufort, NC, USA. He was born in Southampton, England and educated in Canada. He received his Ph.D. from the University of Guelph in 1990 for research conducted on the life history and bycatch of harbour porpoises in the Bay of Fundy, working under the supervision of Dr. David Gaskin. He has conducted field research on marine mammals, sea birds and sea turtles in North and South America, Europe, Asia and the Antarctic. Andy is active in the conservation of marine vertebrates at the national and international levels. He has acted as a member of the Cetacean Specialist Group of the IUCN, the Scientific Committee of the International Whaling Commission, the International Committee for the Recovery of the Vaquita and several federal marine mammal Take Reduction Teams. He has served on the Editorial Boards of Marine Mammal Science, the Journal of Cetacean Research and Management and Endangered Species Research. From 2008-2010 he served as President of the Society for Marine Mammalogy. He was recently nominated to serve as Chairman of the U.S. Marine Mammal Commission by President Obama.

Brandon Southall
Dr. Brandon Southall is President and Senior Scientist for Southall Environmental Associates (SEA), Inc. based in Santa Cruz, CA, a Research Associate with the University of California, Santa Cruz (UCSC), and an Adjunct Assistant Professor at Duke University. He obtained Masters and Ph.D. degrees from UCSC in 1998 and 2002, studying communication and hearing in seals and sea lions. From 2004 to 2009, Dr. Southall directed the U.S. National Oceanic and Atmospheric Administration (NOAA) Ocean Acoustics Program, within the National Marine Fisheries Service, Office of Science and Technology. In 2009, Dr. Southall founded SEA, a research and consulting small business conducting and applying science to support conservation management and environmentally-responsible development primarily (see: www.sea-inc.net). Brandon has an extensive technical background in leading both basic and applied laboratory and field research programs as well as applying science in national and international policies. The largest such effort has involved his serving as the chief scientist for a major multi-institutional behavioral response study supported by the U.S. Navy to study marine mammal responses to military sonar systems. He also serves as a technical advisor to international corporations and environmental organizations regarding the impacts of conventional and alternative offshore energy development and commercial shipping. He has published over 60 peer-reviewed scientific papers and technical reports, and has given hundreds of presentations on related subjects to scientific, regulatory, Congressional, and general public audiences around the world.

Len Thomas
Dr. Thomas is an ecological statistician at the University of St. Andrews. He is the director of the Centre for Research into Ecological and Environmental Modelling and a reader in the School of Mathematics and Statistics. He is also a member of the UK National Centre for Statistical Ecology and the Scottish Oceans Institute. His main research areas focus on the development of methods and software for estimating the size, density, and distribution of

wild animal and plant populations, and the use of computer-intensive methods to fit and compare stochastic models of wildlife population dynamics and animal movement. Of relevance to this project, he has led research projects developing methods for quantifying marine mammal density, distribution and trends (particularly from passive acoustic data), analyzing cetacean behavioral response studies and quantifying the population consequences of anthropogenic disturbance. He has also served on the BP-sponsored Working Group on Assessment of Cumulative Effects of Anthropogenic Underwater Sound, as well as ONR's Population Consequences of Disturbance Working Group. Dr. Thomas received his Ph.D. in Forestry from the University of British Columbia

<u>Peter Tyack</u>
Peter Tyack a behavioral ecologist who studies acoustic communication and social behavior in marine mammals. He has studied reproductive advertisement in baleen whales, individually distinctive contact calls, and echolocation in deep diving toothed whales and has developed new methods to sample behavior continuously from marine mammals, including the development with engineer Mark Johnson of sound-and-orientation recording tags. He has developed a series of studies on responses of marine mammals to anthropogenic sounds, including effects of oil exploration on baleen and sperm whales, and the effects of naval sonar on toothed whales. Peter has extensive experience advising non-governmental groups and government agencies on effects of anthropogenic sound on marine mammals, and is an author of 3 reports on the effects of sound on marine mammals published by the National Academy Press.

Appendix 4 - Submitted Reviewer Biographies

Vincent Janik

Vincent M. Janik, Ph.D., is a Professor at the School of Biology of the University of St Andrews in the UK. Before joining the St Andrews faculty, he was a research fellow at the Woods Hole Oceanographic Institution, USA and a Royal Society University Research Fellow at the University of St Andrews, UK. In his research, he concentrates on vocal communication and the effects of noise in marine mammals, and the evolution of complexity in animal communication and cognition in general. He has published numerous research articles on these topics in scientific journals such as *Science*, *PNAS*, and *Current Biology* and holds patents on acoustic deterrence techniques for mammals. In 2003 and 2009, he was invited as a fellow to the Centers for Advanced Study in Berlin and in Budapest. Currently, he serves on the editorial boards of the journals *Behavioral Ecology and Sociobiology* and *Animal Cognition*. He is also the main editor of the Springer book series *Animal Signals and Communication*.

Hans Slabbekoorn

Hans Slabbekoorn is Associate Professor at Leiden University, the Netherlands. He is specialized in the acoustic ecology of birds and fish with outdoor and indoor work on fundamental and applied aspects. He has published over 65 papers in peer-reviewed journals and his current h-index is 25. He is a board member of the "Dutch Association for Behavioural Biology" (2010-); principal investigator in the NWO-ZKO project on the "Effects of Underwater Noise on Fish and Marine Mammals in the North Sea" (2011-); associate editor for the international journal of "Evolutionary Ecology" (2012-); advisory board member for the conferences on the "Effects of Noise on Aquatic Life" (2012-); advisory board member for the EU-SONIC project on "Suppression of Underwater Noise by Cavitation" (2013-); and project leader for the JIP research project PCAD4Cod on "Population Consequences of Acoustic Disturbance for Fish" (2015-).

Douglas Wartzok

Douglas Wartzok is Provost Emeritus and Professor of Biology at Florida International University. He received a B.A. in Physics and Mathematics from Andrews University, a M.S. in Physics from the University of Illinois, and a Ph.D. in Biophysics (Neurophysiology) from the Johns Hopkins University. He has been a faculty member and academic administrator at Johns Hopkins University, Purdue University, University of Missouri-St. Louis, and Florida International University.

His research on marine mammals has taken him from the Arctic Ocean to Antarctica to study seals, whales and walrus. He along with his colleagues and graduate students have developed acoustic tracking systems for studying polar seals under the ice, and radio and satellite tracking systems for studying whales. His research focuses on behavioral and physiological ecology of marine mammals; sensory systems involved in under-ice navigation by seals; and psychophysiological studies of captive marine mammals. For the past decade he has been involved in the issue of the effects of naval anti-submarine warfare sonar on marine mammals, in particular beaked whales.

For eight years he edited ***Marine Mammal Science*** and is now Editor Emeritus. He served as Chairman of the Committee of Scientific Advisors, U.S. Marine Mammal Commission. He is the Chair of the Committee of Scientific Advisors of the Society for Marine Mammalogy.

He was a member of the National Academy of Sciences Committee on “Assessing Ambient Noise in the Ocean with Regard to Potential Impacts on Marine Mammals,” and chaired the National Academy of Sciences Committee on “Determining Biological Significance of Marine Mammal Responses to Ocean Noise.” He is a member of the National Academy of Sciences Ocean Studies Board and is a member of the NAS Committee on Cumulative Effects of Human Activities on Marine Mammal Populations.

Appendix 5 – Submitted Reviewer Reports

Vincent Janik

To mitigate environmental impacts, the US Navy has a long standing interest in behavioural responses of marine mammals to acoustic sources that they introduce into the marine environment for training and operational purposes. Their interest has led to a comprehensive research program funded by the NAVY on this subject. This document presents my evaluation of the outcomes of this program and recommendations for future work based on the published literature, discussions with PIs and the Behavioral Response Research Evaluation Workshop held in Monterey in April 2015. The research program has two main approaches, one being observational focussing on animals around Navy operations, the other experimental using playbacks of treatment and control sounds to test hypotheses and establish dose-response functions. In this review, I will not summarise findings but concentrate on the next key questions and recommendations for future work.

Comment on the work conducted so far

The Navy program has funded many excellent studies on the topic of responses to sonar. Initial studies on beaked whales have shown interesting reactions to sonar, either using passive acoustic monitoring (e.g. in the AUTEC range) or DTAGs to record responses on the animal. Both showed interesting patterns and contributed greatly to our understanding of beaked whale biology. A large part of this was necessary methods development to study these animals. Further development is needed in attachment methods and duration of DTAGs to achieve the necessary sample sizes for these species. Studies on more accessible species have contributed a tremendous amount to our understanding of cetacean reactions to sound. Species like pilot whales, blue whales, and killer whales were more accessible and therefore studies resulted in larger sample sizes. Passive acoustic monitoring studies supplementing DTAG work showed interesting reactions but are limited by their dependence on vocally active animals. Captive work has been used to test tolerance and cooperation of trained animals in training contexts, but there are doubts whether these data represent tolerance in wild, untrained animals. More could be done here by exposing captive animals unexpectedly to sound and by using choice experiments to determine preferences and aversion to acoustic stimuli. The focus on the DTAG has perhaps also decreased observational efforts and the group focussed sampling methods for surface observations seem perhaps somewhat unsuitable to evaluate reactions at the level of detail that is required. More efforts to expose animals while at the surface combined with observational methods that do not require tag attachment but follow individual animals could be used to increase sample size in beaked whales and other species. Overall, however, the program has to be complimented on a tremendously successful effort to further our understanding of how cetaceans react to sonar. The following comments are therefore not to be seen as a criticism of the existing program but as recommendations to continue its success and address the questions that have emerged from the data that have been collected so far.

Future key questions and recommendations for further studies

What is the mechanism behind behavioural responses?

Stranding events observed around some Navy exercises are clearly not adaptive. It is difficult to see how an animal that shows fast spatial avoidance in response to a signal as shown in some BRS studies would eventually swim onto a beach. This destructive response suggests that there might be two different mechanisms as work, one that explains the behaviour at low levels, while a different one such as the sensitization to startling stimuli as demonstrated by Goetz & Janik (2011) may explain stranding events. An exposure to higher levels than used in BRS so far would be desirable here, either in an experimental setting or by monitoring animals around an actual Navy exercise where high levels are more prevalent. It would allow us to see whether more extreme reactions develop out of low level avoidance, or whether there is a sudden shift in reactions at a particular threshold. Understanding what mechanisms are at work is the key to being able to predict reactions when operating around marine mammals. It will allow to produce informed guidelines when and how to use sonar around marine mammals and how to avoid lethal events. BRS studies conducted so far all assumed one mechanism that explains all behavioural responses with extreme events occurring when the signal exceeds received levels above natural variation. The best example is the assumption that the reaction represents predator avoidance which goes to an extreme if the threat appears louder than normally possible. This is focussing on the received level, but reactions may be caused or modified by other parameters. A startle response, for example, only occurs when a received level is supplemented by a short onset time. A reaction to a startle stimulus at low levels, (i.e. less than 90 dB above the hearing threshold) may be habituation, while the reaction when exceeding this level turns in to the opposite when animals are sensitizing to it. At low levels, received level might be responsible for the reaction, but at high levels it is the combination of received level and onset time. While received level is clearly a very important aspect, more parameters need to be tested for their potential to modulate responses. Captive choice studies can be useful here in which animals are presented with different sounds in different parts or pools of their enclosure to observe where they choose to spend more time indicating a greater tolerance for the associated signal.

What is the relationship between source distance and received level?

From past studies it appears that animals react differently to sources that are close than those further away, even when the received level is comparable. Many animals use sound deterioration as an indicator of the distance to the sound source. It is likely that marine mammals do the same. Since many signals are tested with scaled sources it is important to understand how received level and distance interact when a behavioural response occurs or is absent.

Which species are most vulnerable and how do they react?

While this is a question that has been addressed by several BRS studies, sample sizes for the most vulnerable species are generally too low. While these may be sufficient for a particular statistical method, they are too low to represent natural variation in an adequate way. Every effort should be made to increase sample sizes for these studies. Notable exceptions are efforts on blue whales and other more easily tagged species. However, the reactions in these species appear to be subtle and do not resemble the sometimes extreme reactions of beaked whales for example. While it is generally accepted that beaked whales are of particular concern, many species remain untested. Small delphinids are the most frequently exposed animals in Navy exercises. Research on these animals should be a high priority given the number exposed each year.

What are long term effects of noise exposure?

Repeated exposure may lead to increased stress and ultimately to fitness consequences. New approaches need to be explored to investigate such effects and to assess when sound exposure has the potential to become a problem on the population level. Studies addressing these questions would look at the effects of noise on stress levels, and how stress modifies behavioural responses. Body condition studies may allow quantifying long term changes when comparing animals at the start and the end of an exposure period in a chosen location, or in observational studies when comparing body condition in exposure zones such as on the Navy ranges to those of animals that live in areas with little sound exposure. A change in body condition is clearly not a behavioural response, but would lead to behavioural modifications due to the resulting changes in energetic requirements. By investigating these relationships, we may be able to predict long term effects by observing initial responses to noise exposure.

Are there cumulative effects in noise exposure?

Cumulative effects may explain some of the more dramatic reactions to noise. These can take various forms. Repeated exposure has been tested in many BRS studies by exposing the animals over a period of time. What is lacking are studies with two sources, or repeated exposure of the same individuals over several days. Similarly, different source movements may lead to different reactions. Sometimes animals may get chased without operators being aware of this happening. This may lead to a more pronounced effect. The 3S project has applied such an approach, but did not compare it sufficiently to non-directed passes.

The role of experience

Comparisons of animals that live in areas with frequent Navy exercises and those that live in quieter environments may help to understand the role of experience in noise exposure. Similarly, one could compare the reactions of young and old animals to the same stimuli or conduct longitudinal studies on the same individuals in resident populations. In elephants, experience has a major influence on how a group reacts to an acoustic challenge (McComb et al. 2001). A similar role of old animals in assessing threats is likely to be found in cetaceans.

Contextual aspects of reactions to noise

Context clearly modifies how animals respond to noise. This has been demonstrated already in case studies from existing BRS studies. It is possible that animals suffer long-term consequences when exposed to noise, but remain in an area because it is a lucrative foraging ground (e.g. Hastie et al 2015).Ultimately, these animals may still be compromised by the exposure to noise. Studying the contexts in which animals react or do not react is therefore crucial to understand the impacts of sonar.

Usefulness of approaches

A variety of approaches has been used in the program so far. Several of the available methods complement each other and such combinations should continue to be of value to the program. The development of the DTAG is perhaps one of the most influential developments of the Navy research program. The usefulness of this tool should not be underestimated, not only for sonar studies but for marine mammalogy in general. It allows controlled experiments on individuals in the wild, and this is and will remain one of the most useful ways of studying responses to noise. The experimental approach that can be used when tagging animals allows us to compare the reaction to treatment stimuli in comparison to control stimuli. This comparison provides a scientifically strong assessment, which cannot be achieved by comparing reactions to sound sources with baseline behaviour. Captive studies are often criticised for not being representative for behaviour in the wild. However, they have a role in testing directions of responses that can then be further studied in wild animals. For example, general aversiveness of parameter combinations can be tested in captive animals. The initial reaction to such novel stimuli will be similar to those in wild animals and will allow to decide what to test further in the wild. This is a much more efficient way to narrow down the parameters of interest. Similarly, new methods can be tested on captive animals to save time and money when deploying them in the wild.

Passive acoustic monitoring studies allow detection of vocalizing animals and can be useful to monitor marine mammal presence in areas of interest. Ideally, PAM should use an array that allows tracking or some other way of assessing the number of animals vocalizing beyond just counting vocalizations. Early studies have shown that vocal activity may stop in response to noise exposure. This can either be caused by animals moving away or by stopping to vocalise. Data from tagged animals will allow us to interpret such reactions. Once it is established what explanation accounts for vocalizations to stop, PAM may be used to look at reactions in animals that normally vocalize continuously. When large arrays are available in areas with resident populations, individual tracking may be possible as well using individually distinctive parameters in vocalizations.

Observational data can be collected from fixed observation points or during focal follows when following focal animals. The latter method has been underutilised and should be used to provide more information in the future. Observations from fixed points seem less useful for studying reactions except for initial responses. All behavioural follows should focus on individuals since group sampling is highly problematic (Mann 1999).

A combination of methods as has been used in the AUTEC studies is often helpful to maximise our understanding. However, efficiency has to be considered. Tagging animals in areas where real Navy sources are about to operate does not seem very lucrative since tagged animals are likely to move on before sound exposure starts. Surface observations may be more helpful in these situations.

A technological development that would be helpful in this study program is a longer term attachment of an acoustic tag. Such a tag does not necessarily need to record full bandwidth audio continuously. A tag that processes acoustic input and provides a cumulative or count measure of sound exposure would be helpful to look at longer term and cumulative effects. Data transmission could use satellite or radio techniques if sound data are sufficiently processed and summarised on the tag. In this section, I should also point out that my qualification does not allow me to evaluate the best possible statistical approaches. While the empirical data are collected by a range of research groups and PIs, the statistical treatment of data appears to be combined in only one group. A review of statistical methods by an independent reviewer would therefore perhaps be advisable.

The importance of baseline studies and long-term monitoring

All available data make it apparent that we still have too little knowledge of baseline behaviour in a lot of species that are of interest. Baseline data help to understand natural variation in vocalization rates, movement patterns and dive behaviour. This knowledge is crucial for the interpretation of reactions to sound stimuli. A focus on baseline data is therefore advisable. Similarly, long term monitoring of individuals after exposure should be a priority to understand the time window which is affected by sound exposure. In birds, a brief playback of an intruding male can affect singing behaviour over up to24 hours (Amrhein&Erne 2006). This kind of information is important to have when studying the effects of noise, and will allow predicting longer-term and population level effects more clearly than the short term observations that are available now. Existing satellite tag work has started to look at this time line, but need to be augmented with new tags that can measure sound exposure.

Literature

Amrhein, V. & Erne, N. 2006 Dawn singing reflects past territorial challenges in the winter wren. *Animal Behaviour* **71,** 1075-1080.
Götz, T. & Janik, V.M. 2011 Repeated elicitation of the acoustic startle reflex leads to sensitisation in subsequent avoidance behaviour and induces fear conditioning. *BMC Neuroscience* **12,** 30.
Hastie, G.D., Russell, D.J.F., McConnell, B., Moss, S., Thompson, D. & Janik, V.M. 2015 Sound exposure in harbour seals during the installation of an offshore wind farm: predictions of auditory damage. *Journal of Applied Ecology* **52,** 631-640.
Mann, J. 1999 Behavioral sampling methods for cetaceans: a review and critique. *Mar. Mammal Sci.* **15,** 102-122.
McComb, K., Moss, C., Durant, S.M., Baker, L. & Sayialel, S. 2001 Matriarchs as repositories of social knowledge in African elephants. *Science* **292,** 491-494.

Hans Slabbekoorn

During the meeting there was agreement from the sponsors that they are looking for a conceptual overview and a review of approaches rather than specific projects. They would like to see the strengths and weaknesses of each approach assessed and for different integration opportunities to be rated. Therefore, I did the following: 1) I reviewed the remarks and suggestions made by others in the questionnaire and picked out those that I wanted to emphasize or for which I thought it may be useful to add some additional comments. 2) I inserted notes that were made during the meeting on what I thought were important issues or insights. And 3) I added some additional issues that I came across while reading the literature or thoughts that I had after the meeting.
I followed the structure of 6 "research topics" as used for both the questionnaire and the meeting. The first research topic concerned "General overview questions", while the second until the fifth concerned different classes of methodology: captive studies, scaled exposure, real exposure, passive acoustics. These four classes reflect a series of useful approaches with a step-wise decline in the level of experimental control and a stepwise incline in the degree of realism. The sixt research topic concerned the "Interpretation of responses", in which the use of playback of predator sounds were evaluated.

Research topic 1: General overview questions.

There were a number of questions that stood out to me as clearly of key importance or accurately reflecting what should be the main goal. In most general terms it seems that this is the main question: "What actions can the Navy take to avoid any future sonar-linked strandings, whilst minimizing impacts on their operational and training activities?", certainly if extended to any other impact of significance reflected in this question: "What are the long-term population effects of sonar operation on a given species or local stock?"

In order to get towards the answers to these questions there are a number of critical steps to take and issues to figure out that are recognized and reflected best in my view by the following questions: "What are the acoustic features of the signals that trigger pronounced responses, e.g. frequency modulation, time-frequency bandwidth product? What is the most relevant dose term and how can we extrapolate findings to new species? Is dose comprised of sound intensity, distance from sonar to animal, and/or some other metrics? How are responses modulated by behavioral state or other intrinsic factors? How does biological context modulate the response to sonar exposure? Does prey respond to sonar and is that affecting the behaviour of the predator? Under what circumstances do such responses have consequences for the fitness of affected animals? Do short-term responses, such as brief changes in behavioral state, have long-term consequences for individuals or populations?"

As a consequence of the key questions identified above, it is clear that it is very important to "Understand the baseline ecology and energetic costs of individual variability among behavioral responses. This will require development of approaches that link behavioural response to fitness through the currency of energetics." I believe all approaches used so far have yielded useful data and valuable insights and clearly have complementary potential. Each method has limitations and special assets. Therefore, future investigations are likely best "a combination of experimentation and observation that enable linkage of short-term behavioural response to long-term fitness consequences of repeated exposure." An explicit goal and requirement for the interpretation of results from these studies is to "understand the acoustic world of animals including the natural use of sounds for communication and orientation."

Another important issue that also deserves special attention and that will be addressed in more detail again below is the fact that: "No two sites are identical." The remark made about "Minimizing and understanding the differences is vitally important" reflects in my words the critical requirement for any study that the sample population should represent the population for which we raised the question in the first place or for which we want to make a statement. This means that if we like to make a statement about all individuals of a particular species of beaked whales, some of which live in areas that are often exposed to sonar and some of which live in other areas that are less exposed, that investigations in just one particular area do not suffice.

Research topic 2: Responses to *simulated sources* of Navy sonar – Controlled Exposure Experiment (CEE) studies on *captive animals*

In this section, I was surprised by a number of bold statements and the apparently more variable range of opinions among the researchers. This is probably most clear in this statement: "Using captive studies is a waste of time. The results are suspect, especially as they relate to behavioural response." I disagree with this statement, although it is clear that dose-response studies for behavioural changes should not be translated from captive animals tested indoors to wild-ranging animals exposed outdoors. However, in the same way, any study is suspect that is not sampling a population that is well-representing the animals and conditions of the target population for risk assessment. As sample sizes are typically limited and restricted to particular areas, seasons and test conditions, this should always be a major worry for sonar-impact studies, not just for captive studies.

The same issue with captive studies was also phrased in a different way in that they would not be "ecologically valid for studying behavioural responses or stress. They induce potential bias because the test subjects are anticipating reward for participation so may not be representative of the reactions of wild naive individuals." Again, I disagree with the statement, although it is correct to question the validity of extrapolation of absolute thresholds from captive studies to wild-ranging animals in outdoor conditions. However, I believe captive studies can be useful to get significant insights, especially if we accept the fact that dose-response patterns will vary for example by species, region, season, prior experience, or group size and absolute threshold criteria should not be regarded as the single-most-important target of study.

Behavioural effects occur within the audible range of a species determined by detection thresholds at the low-end and more or less by physical damage thresholds at the high-end. There is growing awareness that it is critical to study variation within these broad limits and to gain insight into the many factors that determine whether or not a behavioural effect occurs and whether or not there is a detrimental effect. Several factors can be investigated best with optimal exposure control and knowledge about individual characteristics and experience. Consequently, captive studies could for example contribute to our understanding of the effects of temporal patterning, the potential for habituation and sensitization and the role of individual personality.

Personality refers to variation among individuals of the same species reflected in different response tendencies that are consistent across time and space. Variability in individual response tendencies may mask or bias our interpretation of general behavioural response patterns, especially in the case of small sample sizes. One may argue that typical response patterns in the field are determined and assessed at group level and therefore individual variation will average out and be less important. However, it may also be the other way around in that group responses are determined by the most easily frightened and most responsive individual. If this is the case, it would elevate individual personality to top priority in predicting response thresholds at the group level.

Another potential field of study mentioned at the meeting that could be done in captivity is developing more advanced and more invasive physiological tags. Yet another type of study could be to investigate potentially detrimental effects on performance during natural tasks that are critical for determining vital rates. There is for example growing evidence in a wide variety of taxa that elevated sound levels negatively affect performance in tasks that do not even need to have an auditory component. Responses to visual stimuli can be delayed or scores to cognitive tests can be lower due to noise-induced attention shifts or neural processing efficiency. Captive animals may for example be used to test whether marine mammals that remain in an area and that do not show any conspicuous behavioural response may still be negatively affected by audible sonar if exposure has the potential to lower foraging efficiency.

Research topic 3: Responses to *simulated* Navy sources *or* real but scaled Navy sources on research vessels – CEE studies on *free-ranging animals*

These studies gain in realism relative to the captive studies at the expense of some control over experimental conditions and insight into animal characteristics and individual and group histories. However, there is a general positive attitude to the accomplishments and according to the experts “we have clear understanding of species that respond more uniformly in experimental exposure conditions (e.g. some beaked whales, killer whales) vs. those that show higher individual and behavioral state variability (e.g. blue whales).” The latter refers to “the direct integration of ecological measurements into simulated sonar experiments, which provides direct evidence for the role of foraging context in affecting the probability of response and the consequences of those responses.”

I believe that the suggested pattern of species differences in individual variability may require more tests under more conditions and in different areas on the same species. Furthermore, this theory would also gain external power with data on more species. And obviously, we also need to know what this means in terms of impact; is it more or less harmful to a population if many individuals are affected to a moderate extent or few individuals to a large extent? It will probably be necessary to find out whether variation in response tendencies between and within species are actually related to the detrimental impact of the disturbance. Some species may alter their behaviour quickly, to anthropogenic stimuli as well as to many natural stimuli throughout the day, and even a strong modification of behaviour by sonar may not concern a significant disturbance that would be measurable energetically. Other species may be less easily affected in changing their swimming direction or foraging activity and for them a subtle behavioural change in response to sonar may actually have more serious consequences.

Another issue worthwhile to address here was the potential strength of a clearly balanced experimental design with the appropriate controls. One statement reports: “The CEE studies have relied heavily on the need for a 'control'. Control periods are often confounded by any number of variables including incidental sonar, vessel traffic, social interactions, etc., thus further increasing the number of exposures needed to account for these differences. Often, the control period is of such short length that tag attachment duration is not adequate to document fundamental changes to behavioral patterning.” Although it is not so clear what types of control are meant here, I believe I largely agree here. The strength of the tagging and exposure studies is the potential to control properly, which should also be done then.

CEE-studies including tagging and subsequent experimental exposure need to sample control trials in which every aspect of the experiment is done, except for the sonar exposure. Treatment and control trials should be assigned randomly or in a pre-determined balanced way to avoid effects of order, time or day, season, region, etcetera. In that way, any impact of vessel presence and tagging can be controlled for. This type of control is not often taken serious enough and only included to some extent. A group or individual can also be its own control by using baseline data from after tagging and before exposure to compare with their behaviour during and after exposure as long as the tag is still on. However, it was reported that there is “often not sufficient baseline data to put exposures in the context of individual variability.” And I would add that baseline data are often very variable, making it hard to reveal any significant change in behaviour at the on-set and during the sonar exposure. An improvement may be to get longer periods of tagging and be more selective in the periods that are used as baseline, really matching the particular environmental conditions and behavioural mode the animals were in at the moment of exposure.

A matter of interest that was heavily discussed during the meeting was the role of distance and level in determining effect size. There were strong expectations that the same receiver level may yield different response tendencies based on whether the sounds were coming from a moderate source nearby or a louder source further away. The expectations were in my perception based on very few observations that probably varied in more parameters and this phenomenon may concern a misinterpretation of anecdotal reports with unknown confounding variables. However, on the other hand, we do know that birds recognize degradation-related properties of sounds transmitted through their typical environment.

Birds are able to range songs of competitors by amplitude as well as other degradation-related acoustic features. Conspecific songs that are of equal amplitude at the receiving bird, but that concern playbacks of recordings at short and long distances trigger short and long response flights respectively (leading to some birds overflying the playback speaker due to erroneous distance assessments beyond). As far as I know such ranging abilities have not been shown in any species for deterrent sound sources that would induce variation in fleeing response tendencies due to distance assessment not based on amplitude but on other degradation and propagation-distance related acoustic features.

Research topic 4: Responses to *real* Navy sources deployed by Navy vessels – *free-ranging* CEE studies

Also for this type of study it was noted that "distance between animal and source is a critical factor in producing a behavioural response." It was therefore noted that it is important to "improve our understanding of how changes in behavior scale with range-to-source, particularly for distant/low-received-intensity exposures." In another statement it was phrased in this way: We need to break the correlation between received level and source distance to determine the relative importance of proximity and received level. Changing the source level (if possible) will help. Using two different sites with significantly different propagation loss so that the correlation between received level and distance is significantly different between the two sites. Responses may occur at much larger distances than for simulated sonar and it will be difficult to cover this in an experiment". I agree that experimental tests are possible to solve the issue, but the suggestion provided here will always be flawed due to potentially confounding variables at the two sites.

I believe the challenge of an(y) experiment addressing the issue above would be the sample size and distance control between animals and sound source at sea. However, in term of experimental design, it would be not too difficult to tease these factors apart in an ideal world. For example, one could test response tendencies for a replicate set of sonar sounds that are recorded at two different distances and played back from the same distance at the same normalized level (Degradation different/Receiver level same). One could also test response tendencies to a replicate set of sonar recordings that are played back at two different levels from the very same distance (Degradation same/Receiver level different). Each test would be strongest in a paired design if it were possible to test individuals or groups twice to get responses to two exposures. If so, the order should obviously be alternating and taken into account statistically. When I state replicate set here then I mean that every pair of sounds played back to another individual or group ideally should be a unique recording. The replicate set of recordings should then be a sample that reflects the natural variation in sonar exposure conditions that we are interested in.

Note that the issue of distance versus level applies to both scaled and real sources, but that the suggested experimental design may be best done with real sonar sounds that are recorded and then played back at the target levels. In addition, the issue of the importance of sufficient and adequate baseline data also applies to both scaled and real sources. It may be interesting to mention that zebrafish studies in captivity and seabass studies in captivity and in large outdoor enclosures also show highly variable baseline patterns between and within groups (even in their simple and restricted captive environments). In general and independent of species, this obviously causes problems in finding significant changes and elevates the required sample size.

Another point mentioned that may yield interesting insights is any advancement in tagging ability that would allow following multiple individuals during the same exposures. As fleeing behaviour has such a strong social component, interactions among individuals may explain why behavioural effects may escalate in some cases or fade out quickly in others. Tagging of multiple individuals may reveal such interactions and also allow the study of individual differences to the very same exposure situation. Using the real thing instead of scaled playbacks adds realism to these types of experiments, but control exposures to other sounds such as white noise or even other man-made sounds, such as pile driving or airguns may allow comparisons in the qualitative and quantitative nature of behavioural responses and put sonar into context. Are the responses sonar-specific or just artificial sound specific?

Research topic 5: Responses to *real* Navy sources and Navy vessels – *observational* (e.g., opportunistic or incidental) studies

The last of methodological classes concerned the one with the least experimental control but also the one with the most realism, as there are no observation vessels or other confounding variables present at the moment of unplanned exposures. "These observational studies have been very important in demonstrating that, at least in some areas (such as Navy ranges), cetaceans are repeatedly exposed to Navy sonar without any apparent immediate harm. The studies have also been important in demonstrating that short- to medium-term displacement is a common response for some populations (and one that could not be demonstrated using experimental approaches using short-term tags). Recording tags typically last for 24 hours." Another quote addressed the same issue of tag limitations: "An increase of data logging capability to even 48 hours would significantly improve the possibility of tagging beaked whales around actual Navy operations. The additional attachment time would also provide extended data that would help to put any reaction in context."

There are also several notes on other short-comings as for example this type of "observational studies lacks control over the exposure conditions. So although they are relevant to actual Navy activities, they may not allow answering fundamental questions regarding animal responsiveness to sound." And also: "The observational methods do not allow direct measurements of received levels and exact source-whale distances are observed relatively infrequently and sometimes with significant uncertainty. The data on animal behaviour is also not as fine-scale. So, in general, these studies focus on displacement of animals relative to a source or exercise, rather than on fine-scale behavioural changes." Observed responses are potentially even just from the less responsive part of the population (more sensitive animals may have left). On a positive site it was also noted by someone that "there is some work in progress where dive profile data are available in addition to animal position data, which would be an improvement " in terms of the potentially unique value for this type of methodology.

In my view passive recordings certainly add value to the marine mammal studies tool box. "These methods (both satellite tagging and PAM) may be useful in identifying general patterns of activity (e.g. movement, basic aspects of diving) and how they may change in periods with and without sonar. An important limitation is that we cannot distinguish between two very different types of response – a reduction in vocal activity versus physical displacement outside the range of the receiver." Also explicitly mentioned by someone in the questionnaire is that "these methods in deliberate combination with experimental methods can provide useful linkages from individual responses to population level estimates." And I agree as especially the long-term nature of some data-sets may open up possibilities to fill in knowledge gaps for a Population Consequences of Acoustic Disturbance model approach. "The longer term measurements can allow bioenergetics analyses based on patterns of typical movement" and in this way passive acoustics data are for example already being used to inform the development of a beaked whale PCAD model".

A PCAD model approach doe not only require insight into natural and modified patterns of swimming and diving, but also on whether foraging opportunities have changed due to these altered patterns and whether foraging efficiency is affected in any way. It was mentioned that "Displaced animals move to alternate areas to forage. Data on the nature of the prey field is lacking especially for deep divers such as beaked whales." However, I have seen very few papers in which we have detailed knowledge on prey abundance and quality in the area that was left behind as well as in the alternate area. Furthermore, prey may also respond to sonar and alter water column depth or swimming patterns and the presence of sound may also alter the foraging efficiency. The consequences of sonar exposure events will have immediate effects as well as lasting spatial effects on future travelling routes depending on the degree of anxiety by the event itself and the altered experience in terms of feeding rates and quality.

Research Topic 6: Interpretation of responses

Interruption and deterrent effects of sonar exposure are typically interpreted as anxiety-related and attributed to a perceived threat of elevated predation risk. This is despite the fact that very few encounters with sonar-producing sound sources will increase mortality in any marine mammal species. However, artificial and unfamiliar sounds or any sign of human presence may be enough to elicit anxiety. In line with this theory "an anti-predator response is the leading hypothesis proposed to explain the adverse response exhibited by some cetaceans to Naval sonars." Consequently, it has been argued that "comparisons with anti-predator responses to other natural threats as a template is a powerful approach to address the perceived predation risk hypothesis."

Several problems are also identified for the predator playback studies (playing back killer whale sounds) along the same lines as for the scaled exposure studies mentioned above. It is for example mentioned that "since both predator response and sonar response are probably highly context-specific, it may be tough and time consuming to achieve an adequate sample size to really understand how predator response relates to sonar response." Nevertheless, there seem to be a general agreement that "by knowing species predator responses one could potentially use that as a measure to relate to those observed during navy training. If species predator responses are consistent one might be able to determine when they are in that behavioral state during contolled exposure studies.

I believe the killerwhale playbacks are valuable and the hypothesis may turn out to be valid, but it remains the issue what the consequences are to the animal. It may be a valuable strategy if the target is to find out what the short-term and long-term consequences are of the reponse in terms of energetic expenditure, stress physiological costs, lost foraging opportunities or social separations. However, it is not yet clear to me whether we gain any insight into the costs or detrimental impact of behavioural effects if there is or is not a match to behavioural patterns under naturally risky circumstances. It may for example be, as mentioned before, that prey species that often shift rapidly in behavioural patterns, due to common events of the potential approach of predators, hardly suffer from doing that once or twice more due to sonar exposure. Other species that are less vulnerable naturally may experience more detrimental impact due to the same or less dramatic behavioural effects.

The discussion above stresses the importance of understanding the general ecology of marine mammals and especially the role of sounds in their life. Interpretation of behavioural effects requires a thorough understanding of physiological and ecological consequences, which rely heavily on knowledge about what the animals typically do and would have done otherwise in the absence of sonar and in the absence of human researchers. In line with these insights there are several remarks made such as: “Any research study subjecting a population to controlled exposure experiments should consider deployment of long-term behavior logging tags outside the experimental context to provide more robust baseline data on the range of behaviors seen in local individuals as they range through their environment” and “Critical data include details of energetics of foraging and locomotion, how animals respond to threats, and whether responses such as social defense against predation may reduce acute risk of injury or death from flight responses.” It should indeed be clear that baseline ecological studies of multiple species are key to our general understanding and the correct interpretation of both behavioural patterns and experimental results.

The problem of understanding behavioural patterns and especially event-related deviations is not trivial in natural environments. One can maybe learn from experience with captive fish studies that can be observed continuously before, during and after sound exposure and for which there are also data on how they respond to visual threats from above, the visual presence of an actual prdator fish, or alarm pheromone in the water that is indicative of a nearby and active predator. Zebrafish studies have shown for example significant changes in behaviour in terms of swimming speed and alteration of swimming height to moderate sound levels that are not interpreted as being immediately detrimental (other than causing maybe brief increases in energy expenditure).

These zebrafish responses to moderate sound levels do not seem to reflect anxiety but appear rather explorative in nature. At higher levels, strong and repeated startle responses, erratic swimming behaviour, speeding and down-ward dives to the bottom, and freezing bouts can be induced, which are interpreted as an indication of anxiety and are considered as potentially harmful through effects on energy intake and expenditure, stress physiology and extrapolated consequences under natural conditions for increased predation risk. However, habituation as reflected in the fading of behavioural response measurements is often very rapid for this species in these conditions.

Note that sound fields in fish tanks are complex and both sound pressure and sound velocity are important to auditory perception in fish and vary considerably and unlike the physical predictability of outdoor conditions along the swimming trajectory of the fish. Crude qualifications like low, moderate, and high are therefore the best way to describe exposure levels in fish tanks for fish and determination of detailed dose-response functions are problematic in fish tanks.

Douglas Wartzok

OVERVIEW OF PROGRESS TO DATE

Overall there has been excellent progress in a number of areas that provide greater understanding of cetacean responses to naval sonar and the potential consequences. The careful stepwise approach to the controlled exposure experiments (CEE) has built the confidence of researchers in, and mollified the concerns of most of the skeptics to, the CEE approach. The results of most of the studies have been consistent with the hypothesis that the responding animals perceive the sonar source as a potential predator. For most species the responses range from a cessation of vocalization to an interruption of foraging to leaving the area of the stimulus. Enhanced alertness, interruption of foraging and exiting from the area are normal responses to a predator. The response of pilot whales, while different from those of most species by approaching rather than leaving the area of the stimulus, does match their normal predator response which is to "mob" a killer whale predator. The most important outstanding questions relate to whether these normal predator responses will have individual and population-level consequences because of their frequency, intensity or cumulative effects.

Early CEE studies showed that the response of tagged beaked whales was to break off foraging dives, ascend to a mid-depth, pause, and then ascend slowly while increasing the distance between the whale and the source. This was an important finding because it disabused a notion in that literature that the negative sequelae in beaked whale responses to sonar were the result of decompression sickness (DCS) triggered by a rapid ascent.

Reviewing dive profiles with and without controlled exposures showed that in most cases the responses of the cetaceans to sonar would not lead to nitrogen tensions resulting in DCS. The theoretical work on compartmental distribution of nitrogen during a dive showed that in the animals where there was a possibility of DCS the critical factor was not their deep dives or ascents therefrom but rather an interplay between the time spent at moderate depths (30-200 m) during which nitrogen loading could occur compared to time spent in repetitive shallow dives (<30 m) where nitrogen could be safely flushed through alternating cycles of decompression and recompression.

The early work was designed to address directly the response of beaked whales to sonar because the stranding data indicated that these species were the most sensitive (Cuvier's beaked whale, *Ziphiuscavirostris*, constitutes 69% of all mid-frequency active sonar strandings). It soon became apparent that the most sensitive species were also the most difficult to study. Hence the overall study design was rethought and the focus broadened to embrace other species that were more amenable to tagging and study but that also were likely to be affected by sonar, just not in as dramatic a manner. A second important new direction was to redesign the sonar simulation source so that it could be deployed from smaller vessels at a variety of locations. The wisdom of both of these redesigns has become evident through the substantial increase in CEE subjects.

Passive acoustics, particularly the passive acoustics associated with tracking beaked whales on naval testing ranges, has provided important data showing that the whales move off the range during anti-submarine warfare exercises but that when the exercises end, the whales return to the range. These studies have shown what is probably the most definitive evidence of potential long term impact of sonar activities by causing substantial temporal and spatial displacement of the animals. Nonetheless all Navy test ranges—Atlantic, Southern California, and Pacific—have resident populations of beaked whales and other marine mammals. Long term demographic studies of range populations compared with less disturbed populations may provide evidence of population level impacts of repeated sonar exposure. They could also address the hypothesis that range populations act as a sink drawing in healthy individuals from surrounding populations.

Captive studies have complemented free ranging animal studies. They have provided dose-response data from sensitive harbor porpoise that are too small to tag with current data loggers. In other cases they have been able to increase the sample size and obtain dose-response curves with multiple animals exposed to a given signal level. These studies have helped to understand some of the observed variability in response in field studies with different shaped dose-response curves (symmetrical and asymmetrical), different evidence of habituation, and different age-specific sensitivities across different species.

The difficulty in drawing conclusions from data that are highly variable based on largely unknown intrinsic and extrinsic factors has required the development of new analytical techniques and the first application the latest statistical techniques to marine mammal behavioral responses. These techniques such as Mahalanobis distance, hidden state models, recurrent event survival analysis, and Bayesian hierarchical models can have broad applicability beyond the current sonar response studies.

Most of the studies have shown that some animals start responding at received levels substantially below those requiring mitigation under current regulations. Some studies have shown that animals responding at low received levels to simulated sonar do not respond to significantly higher received levels from actual sonar operations where the naval ship source is much further away than the simulated sonar source. The relationship between range and received level needs further investigation.

All studies have shown that responses of animals are highly variable based on unknown intrinsic and extrinsic factors. There are questions regarding the applicability of dose-response curves created with one species under one set of conditions to other species and situations. One of the contributions of the program to date has been the clear evidence of the fallacy of attempting to provide one number for the potential takes in any operation. Clearly the operators and regulators need to work with legislative bodies to embed the inherent uncertainty of any take estimates in the regulatory framework.

Overall the progress to date has been outstanding. I am not aware of anything that has been supported that shouldn’t have been supported. A number of different approaches have been supported and each has made valuable contributions. Tantalizing potential answers have been proposed for several of the initially most troublesome issues. The need now is to extend this fine work through increased sample sizes, extended temporal duration and expanded spatial scale including more research in areas where the animals are presumably more naïve than on the naval ranges.

RECOMMENDATIONS FOR FUTURE RESEARCH

There is a plethora of intriguing scientific questions that can be addressed with the multiple tools developed through the supported research to date. My recommendations are based on my view of the most important scientific questions required for Navy stewardship of oceanic marine mammals and responsiveness to regulatory requirements. I think the top three major topics are: enhanced baseline studies, appropriate response indicators, and long-term impact.

Baseline Studies

In order to determine the extent to which the changes in behavior observed thus far are truly atypical, more baseline data are required. The baseline certainly needs to be extended around the time of CEE, both before and after, but much longer periods of baseline data should be obtained both on naval ranges and in relevant species habitats far from naval ranges.

In almost all studies, there has been significant variability in the response of the animals to basically the same received level. It should not be surprising that activities such as social behavior, travelling, foraging, and diurnal and annual cycles will influence the responsiveness of individuals. How important various activities are will vary over the course of a day or year and the time spent in various activities will change over the course of a day or year. Baseline studies are needed to begin to understand the non-stimulus based components of behavior. Data obtained so far indicate that seasonal and inter-annual oceanographic variability impacts behavior and habitat use more than the presence of sonar

Obviously one would like to have baseline data at the same level of resolution as that provided by the D-tags in the CEE studies. However for both the temporal and spatial extent of the desired baseline data, these tags will need to be supplemented by other approaches. Observational studies, particularly in conjunction with PAM, can provide some baseline data but because indications are that the most important responses relate to changes in diving patterns and substantial horizontal displacement, much of the needed baseline data will come from Fastloc GPS satellite tags with the ability to transmit compressed dive profile data.

The NRC report[1] on the biological significance of ocean noise recommended comparing observed responses against the baseline data and determining where on the spectrum of baseline behaviors the observed response fell. As a conservative suggestion for determining biological significance, the report identified the 25th or 75th percentile of normal behavior. For example, it suggested there could be a biologically significant concern if the duration or length of the migration of an exposed animal was greater than the 75th percentile of normal migration time or length. For significant foraging impact, the report suggested the body condition of the animal would be below the 25th percentile of unimpacted animals at an equivalent time in the annual cycle.

Whether or not the suggested percentiles of the NRC report are adopted, it does show the necessity of obtaining sufficient baseline data to make a biologically significant impact determination.

Appropriate Response Indicators

Captive dose-response studies have shown species differences in shape of the curve—symmetric or asymmetric; in age-related sensitivity; and in the occurrence of habituation. There has not been much overlap between the species studied in captivity and the species studied in the open ocean. In one case where the same species, bottlenose dolphin, was studied in captivity and in the wild, the captive studies would suggest that dolphins would show greater behavioral response to sonar than was shown by wild bottlenose dolphins on the PMRF range.

Studies with killer whales have shown that sensation level is a more appropriate metric for measuring dose-response than received level. This suggests that work should continue to obtain audiograms of as many potentially impacted species as possible to generalize from

[1]National Research Council. 2005. Marine Mammal Populations and Ocean Noise: Determining When Noise Causes Biologically Significant Effects .National Academies Press, Washington, DC.

received level to sensation level to probability of response. We still have no audiograms for any baleen whale.

Preliminary data from blue whales suggests that behavioral state is an even more important factor in determining response to a given dose than it is for odontocetes where it is still a significant factor. Across all species the data set linking behavioral state and response needs to be expanded.

When severity of response was considered, there was not a clear dose-response relationship. Higher severity responses could occur at low dose and vice versa. Hence more data are needed here as well in order to begin drawing conclusions regarding mid-term and long-term impacts.

Several studies have shown that both distance and received levels are important predictors of response. This result is not at all surprising if the animals perceive the sonar as a potential predator. Prey typically titrate response against distance and do not flee with the first detection of a predator unless the predator is within range to threaten them. A good deal more data need to be gathered to address this issue. This is an essential issue for translating the CEE into values that can be used for regulatory purposes. It is unlikely that the dose-response functions derived with the simulated sources represent the dose-response functions with exposure to operational sonar where range is an important factor. Obviously the best data will come from CEE in conjunction with Navy ships. Unfortunately, given the difficulties of coordination with Navy ships for CEE much of these data will need to be derived from observation of tagged animals in less controlled encounters.

Along with range, direction of movement of the source can be a significant influence. Other studies of cetacean responses simply to vessel noise have shown greater response to vessels approaching than to those stationary or moving away. The 3S experiments showed that the animals changed direction to always move perpendicular to the vector of the ship's movement.

Long-term impact

Behavioral responses have been demonstrated in all species tested, albeit at variable received levels. Cutting-edge statistics have been developed and refined to demonstrate a statistically significant response in a number of cases. The changes in behavior observed clearly have the potential to have long-term impact as a disruption of foraging is a common response. Displacement is another response that takes energy and may move animals to areas where food resources are diminished. Against this potential however are some broad observations of populations apparently surviving well through decades of naval exercises on the ranges that overlap the home ranges of these populations. Absent the stranding response, are there population consequences of sonar exposure? A first step in answering this question is to obtain data over a longer period of time and over greater spatial extent.

Technological Developments

Tag attachment duration must be improved. The attachment time needs to progress from hours to at least a week and preferably longer. Such an attachment will require either a limpet attachment or some type of fully implanted tag. Because most of the BRS work is being conducted at low latitudes where ARGOS tracking is limited, this tag will need to incorporate

Fastloc GPS. After location, the most important sensor is a depth-time sensor with appropriate processing capabilities to send summarized dive profile information during satellite data transmission. The next important sensor is a hydrophone and associated processing so animal vocalizations and ambient sounds can be summarized and transmitted. I understand that a medium-duration Fastloc location only satellite tag is under development for possible deployment in early 2016.

The longer duration tag can address concerns regarding research effects, particularly those associated with the tagging. An argument can be made that research effects would affect the control and experimental presentations equally but it is much better to remove those effects than to have to make the argument particularly when the extinction coefficient of the research effects is unknown and is likely different for different species and different behavioral states within a species. A longer duration tag will also provide the opportunity to obtain a greater amount of background data to help better understand the range of variation in behaviors of the animals and when observed behaviors represent a significant difference from the normal suite of behaviors under a variety of behavioral situations such as foraging, travelling, and social activity. The longer life tag will also allow more experimental presentations and an extended assessment of post-exposure behaviors.

Passive Acoustic Monitoring

PAM will need to be a component of any long-term studies because it is less expensive and can provide data over longer time spans than even extended duration tags. PAM can provide information on distribution and abundance of species whose vocalizations can be discriminated. It can also track specific individuals in some cases. Obviously it works only when the animals are vocalizing and one of the initial responses to sonar has been to reduce or cease vocal activity. PAM can also be useful for studying dolphins and porpoises too small to be tagged. Algorithms to identify vocalizations to species for these cetaceans are still works in progress. Most of the pelagic delphinids cannot be automatically classified to species based on vocalizations. When combined with visual observations (such as the FLIP studies) baseline data and group responses to sonar and other acoustic disturbance can be observed. Most of these results will be more qualitative than quantitative.

PAM on the Navy ranges has been used to significantly advance our understanding of responses of range animals, particularly beaked whales. Given the depth of feeding and the directionality of beaked whale echolocation clicks, it is unlikely that the necessary data quality could be replicated by a portable PAM operating off the range, but there is still much data that can be collected through continuing PAM operations on the naval ranges.

Energetics and body condition assessment

Energy has proven to be a useful currency in development of the PCAD/PCoD model. It is possible that aspects of body condition can be evaluated through swimming and diving/drifting/sinking/rising data obtained with the accelerometers on the D-tag. Tagging animals in populations regularly exposed to sonar on or close to naval training ranges and comparing similar data with animals in more pristine environments could begin to provide indications of the energetic consequences of sonar exposure, either through increased energy expenditure or reduced foraging success.

Demographic variables

One encouraging finding has been the prevalence of various species of marine mammals, including beaked whales most sensitive to stranding in the presence of sonar, on naval training ranges. However, before this can build confidence in the long term coexistence of marine mammals and sonar, comparative demographics between populations resident on naval ranges and those in more pristine environments are needed to answer the question of whether the populations on naval ranges are a sink for neighboring populations. A recent paper[2] suggests that the sperm whale population in the eastern Caribbean, although increasing in numbers and apparently healthy, is a sink due to probable human-caused mortality in this population. This documentation of a sink population gives much more substance to the hypothetical discussion and emphasizes the need for such a comparison of populations on and off naval training ranges.

Stress

Measurable changes in stress hormones in feces have been associated with corresponding changes ship traffic in the Bay of Fundy[3]. So far little work has been done relative to stress hormones and exposure to sonar but looking at stress hormones could provide information on long-term effects. Fecal stress hormones have been used to track short-term changes in stress. Biopsies have been faulted because of the long-term integration of stress signals but such integration is what is needed for a comparison between animals exposed to sonar on a regular basis and animals living in more pristine conditions. Both fecal stress hormones and stress hormone signatures in skin and blubber should be assessed for possible differences related to exposure to sonar.

Species focus

Because beaked whales are the primary animals with documented individual, if not population, consequences through stranding, I think it is important to continue to place substantial effort in tagging and tracking these animals in spite of the difficulty in conducting such research. The interplay between range and received level and response, including severity of response, should be investigated in beaked whales. Demographic profiles of naval training range populations and other populations can be investigated through visual observations of beaked whales.

ESA listed baleen whales and sperm whales are more amenable to tagging and tracking and because of at least short-term interruption of foraging they also have the potential to experience long-term impacts. These species are appropriate for looking at body condition issues via D-tags.

Harbor porpoise have occasionally stranded and in general show the greatest sensitivity to acoustic disturbance. A lot of good data has been obtained from captive animals but so far nothing from wild animals. A Fastloc GPS tag with a dive profiler that could be attached to the dorsal fin of wild captured harbor porpoises should be a technological design goal. Such a tag would also be very useful in a similar attachment to a variety of oceanic delphinids that

---

[2] Whitehead, H. and S. Gero. 2015. Conflicting rates of increase in the sperm whale population of the eastern Caribbean: positive observed rates do not reflect a healthy population. Endangered Species Research 27: 207–218.

[3] Rolland, R.M. et al. 2012. Evidence that ship noise increases stress in right whales. Proc. R. Soc. B. 279:2363–2368

dominate take numbers because of their population size but for which we have virtually no information on either short-term or long-term impacts. (Note that some of the earliest radio tracking of marine mammals was done on delphinids using a fairly large VHF tag secured to the dorsal fin.) Some delphinids observed passively appear to be more tolerant of acoustic disturbance than the whales studied to date although they do respond when received levels become high.

FUTURE RESEARCH OF LOWER PRIORITY

Prey Response

From what we have already learned indirectly about prey response, it appears to be less sensitive than the marine mammal response. There are several observations of whales experiencing higher received levels before responding when they are feeding than when they are engaged in other activities with the implication that the whales remained because the prey continued to be available. Obviously any sonar effects on prey availability need to be incorporated in the cumulative effects of sonar on marine mammals, but the response of the prey to sonar is likely to be a second order effect at the current level of analysis.

Signal Characteristics

In most cases, it appears that the playback of an actual predator, a killer whale, elicits a response at a lower level and a more sustained response than that to sonar. The sonar has a somewhat lower response level than pseudorandom noise of similar timing and bandwidth although the difference between response thresholds of sonar and pseudorandom noise are not nearly as great as those between killer whale vocalizations and either sonar or pseudorandom noise.

At one point it was thought that CEE would be able to elucidate the characteristics of the sonar signal that trigger responses in marine mammals. Research at this stage seems to indicate that this is not a likely outcome of the CEE experiments. Determining the particular characteristics of the sonar signal that are the most critical for generating a response would be a suitable subject for captive research but there would still be major questions regarding generalization to other species. Species amenable to captive research studies are not the species particularly sensitive in the wild.

Signal characteristics will be another second order effect that, based on evidence to date, will likely have less influence on response than behavioral state and prey availability.

Ramp-up

Calculations based on studies to date indicate that under certain narrowly defined circumstances ramp-up could reduce the exposure levels of animals, provided they responded to the ramp-up appropriately. In most cases the speed of naval ships is such that ramp-up will have little effect on SEL.

Determination of Functional Groups

We still do not have sufficient data to be able to extrapolate from the current set of experimental subjects to broader conclusions about which species may be more sensitive to, or more tolerant of, sonar. Based on the small number of species looked at so far, the classification of cetaceans based on functional hearing groups according the general hearing ranges does not appear to be appropriate for assessing sonar impacts. In the class of mid-frequency cetaceans, bottlenose dolphins (at least on Navy ranges) and pilot whales appear to be relatively tolerant of sonar, with pilot whales attracted to playback sources, whereas sperm and killer whales show increasing levels of response and beaked whales are the most responsive.

When more data are available from a broader range of species, new functional groupings may become more apparent.

REASSESSMENT OF TAKE

While not specifically a research topic, the findings to date with the CEE have demonstrated that NMFS and the Navy need to look at regulatory and possibly legislative changes so that take becomes a more scientifically justifiable concept. Having demonstrated statistically significant behavioral changes occur in some species and for some signals at the level of audibility, the current application of take will result in huge numbers of animals for which monitoring and mitigation would be impossible.

It is true that the "small numbers" requirement of 16 U.S.C. 1371(a)(5)(A) and (a)(5)(D) does not apply to military operations, but still a finding of negligible impact is required in order to provide an IHA.

The Navy needs to make use of the revised Level B harassment definition provided in 16 U.S.C. 1362 (18)(B)(ii) where Level B harassment for military operations is defined as responses that occur "to a point where such behavioral patterns are abandoned or significantly altered."

It is of course subject to interpretation as how long a given behavior (e.g., foraging) is interrupted before meeting the definition of being abandoned. Similarly, significantly altered could be interpreted in a statistically significant sense or in a biologically significant sense. I expect the intent of Congress was more in line with the latter.

The experiments have shown that some form of a dose-response curve, rather than a step function, best describes the impact of sonar on all species studied to date. Thus a scientifically based take would be a number with a confidence interval. The continuation of presenting a single number for the take of any species is not the best available science.