Behavioral responses of individual blue whales (*Balaenoptera musculus*) to mid-frequency military sonar

Brandon L. Southall[1,2], Stacy L. DeRuiter[3], Ari Friedlaender[1,2,4], Alison K. Stimpert[5], Jeremy A. Goldbogen[6], Elliott Hazen[2,7], Caroline Casey[1,2], Selene Fregosi[1,4], David E. Cade[6], Ann N. Allen[8], Catriona M. Harris[9], Greg Schorr[10], David Moretti[11], Shane Guan[12], John Calambokidis[8]

1. Southall Environmental Associates (SEA), Inc., Aptos, CA, USA, 95003
2. Institute of Marine Sciences, University of California, Santa Cruz, Santa Cruz, CA, USA 95064
3. Calvin College, Grand Rapids, MI 49546, USA
4. Hatfield Marine Science Center, Oregon State University, Newport, OR, 97365, USA
5. Moss Landing Marine Laboratories, San Jose State University, Moss Landing CA, 95039, USA
6. Department of Biology, Hopkins Marine Station, Stanford University, Pacific Grove CA 93950, USA
7. NOAA Southwest Fisheries Science Center, Monterey CA, 93940, USA
8. Cascadia Research Collective, Olympia, WA, USA
9. Centre for Research into Ecological and Environmental Modelling, University of St. Andrews, St. Andrews, Scotland, UK
10. Marine Ecology and Telemetry, Seabeck, WA, USA
11. Naval Undersea Warfare Center, Newport RI, USA
12. Office of Protected Resources, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Silver Spring, MD, USA

Corresponding author: Brandon Southall, Brandon.Southall@sea-inc.net

Key Words: Sonar, marine mammal, blue whale, behavioral response, endangered, context

SUMMARY STATEMENT

Controlled exposure experiments using simulated mid-frequency sonar and pseudo-random noise revealed individual variation in behavioral responses of blue whales. Responses depended on contextual factors, including behavioral state, proximity, and prey.

ABSTRACT

This study measured the degree of behavioral responses in blue whales (*Balaenoptera musculus*) to controlled noise exposure off the southern California coast. High-resolution movement and passive acoustic data were obtained from non-invasive archival tags (n=42) while surface positions were obtained with visual focal follows. Controlled exposure experiments (CEEs) were used to obtain direct behavioral measurements before, during, and after simulated and operational military mid-frequency active sonar (MFAS), pseudorandom noise (PRN), and controls (no noise exposure). For a subset of deep-foraging animals (n=21), active acoustic measurements of prey were obtained and used as contextual covariates in response analyses. To investigate potential behavioral changes within individuals as a function of controlled noise exposure conditions, two parallel analyses of time-series data for selected behavioral parameters (e.g., diving, horizontal movement, feeding) were conducted. This included expert scoring of responses according to a specified behavioral severity rating paradigm and quantitative change-point analyses using Mahalanobis distance statistics. Both methods identified clear changes in some conditions. More than 50% of blue whales in deep feeding states responded during CEEs, while no changes in behavior were identified in shallow-feeding blue whales. Overall, responses were generally brief, of low to moderate severity, and highly dependent on exposure context such as behavioral state, source-to-whale horizontal range, and prey availability. Response probability did not follow a simple dose-response model based on received exposure level. These results, in combination with additional analytical methods to investigate

different aspects of potential responses within and among individuals, provide a comprehensive evaluation of how free-ranging blue whales responded to mid-frequency military sonar.

I. INTRODUCTION

Sound production and reception are centrally important in the life history of all marine mammals, and their responses to natural signals as well as human noise can have both positive and negative fitness implications. However, we lack a comprehensive understanding of how most marine mammals respond to sound in their natural environment. Given the substantial scientific and regulatory interest in quantifying the effects of anthropogenic noise on marine mammals in recent decades (National Research Council (NRC), 1994; National Research Council (NRC), 2005; Southall et al., 2007, 2009, 2016; Hatch et al., 2016; National Academies of Sciences, 2017; Southall, 2017), there is a pressing need for detailed measurements of responses to acoustic disturbance in known and/or controlled exposure conditions. Regulatory requirements include quantifying marine mammal behavioral responses to noise with sufficient resolution to understand key aspects of behavior (e.g., foraging) that, if negatively affected, may have fitness consequences at both the individual and population level (King et al., 2015; McHuron et al., 2018; Pirotta et al., 2018).

The effects of military sonars on marine mammals have received particular attention. Specifically, focus has been placed on lethal mass strandings involving beaked whales associated with tactical mid-frequency (nominally 1-10 kHz) active sonar (MFAS) (see: Filadelfo et al., 2009). However, both observational and experimental studies have documented sub-lethal behavioral responses to various kinds of sonar systems in an increasingly wide range of marine mammal taxa (e.g., Fristrup, Hatch, and Clark 2003; Tyack et al., 2011; Miller et al., 2012, 2014; Moretti et al., 2014; Henderson et al. 2014; Sivle et al. 2015, 2016; Isojunno et al., 2016; Southall et al., 2016; Falcone et al., 2017). Responses range from brief and/or minor changes in social, vocal, foraging, and diving behaviors to more severe modifications, including sustained avoidance of important habitat areas in some conditions (see: Southall et al., 2016; Southall, 2017). Although sub-lethal, such responses may negatively influence vital rates in ways that, depending on their duration, severity, and proportion of populations affected, may be consequential for protected or endangered marine mammal species. Direct, empirical measures of sub-lethal behavioral responses of marine mammals are thus needed in contexts where sonar exposure is known and can be compared within and across individuals (Southall et al., 2016). Specifically, given the regular exposure of various species to MFAS in and around military training areas, and the threatened or endangered status of most baleen whale species, understanding the frequency of occurrence and severity of how sonar affects behavior in these species has both scientific and regulatory importance.

Observational studies using passive acoustic monitoring have documented behavioral responses in several baleen whales to various types of operational military sonar systems (Miller et al., 2000; Fristrup, Hatch, and Clark 2003; Martin et al., 2015). Controlled exposure experiments (CEEs) that use high-resolution animal-borne tags with movement and acoustic sensors provide detail on individual behavioral responses as well as the characteristics of received sound at the position of the animal (see: Southall et al., 2016). Such approaches can increase the ability to empirically relate and quantify known sonar exposure with fine-scale aspects of behavioral responses (e.g., foraging) that are more difficult to measure with coarser observational methods. For instance, Nowacek, Johnson, and Tyack (2004) demonstrated responses of some North Atlantic right whales (*Eubalaena glacialis*) to controlled alarm stimuli. Sivle et al. (2016) identified behavioral changes of individual humpback (*Megaptera novaengliae*) and minke (*Balaenoptera acutorostrata*) whales exposed to towed operational military sonars.

Blue whales (*Balaenoptera musculus*) are classified as endangered under the IUCN red list (Cooke, 2018). They are also considered endangered under the U.S. Endangered Species Act of 1973 (16 U.S.C. § 1531 et seq.), which along with the U.S. Marine Mammal Protection Act of 1972 (16 U.S.C. § 1361 et seq.) affords them federal protections within the U.S. Blue whales are the largest animals on the planet, yet they feed almost exclusively on small invertebrates (krill) in near-surface to deep (~300-400 m) layers. They often occur in coastal waters, including along the California coast during summer and autumn. However, they also forage in pelagic areas, including in areas where Navy sonar is regularly used. While, like all baleen whales, there are no direct measurements of hearing in blue whales, they primarily produce and are presumably more sensitive to low frequency sound. However, recent evidence suggests they may be behaviorally sensitive in some conditions to mid-frequency sounds (1-10 kHz).

Behavioral responses of blue whales to MFAS and other mid-frequency sounds have been quantified using CEEs in a series of studies off the southern California coast (Southall et al., 2012; Goldbogen et al., 2013; Friedlaender et al., 2016; DeRuiter et al., 2017). These experimental studies have notably involved MFAS designed to simulate U.S. Navy SQS-53C systems that were used in previous stranding events. The results of this previous work, which involved subsets of the data used here, demonstrate significant behavioral responses of blue whales to MFAS (and pseudorandom noise (PRN), which is of similar frequency and exposure level) across many individuals. Further, they illustrate several context-dependencies in behavioral responses, as noted by Ellison et al. (2012), including strong influences of individual behavioral state at the time of exposure as well as prey distribution and density. DeRuiter et al. (2017) used hidden Markov models to evaluate behavioral state-switching, demonstrating greater probabilities for blue whales to either cease deep-feeding or fail to initiate deep-feeding behavior during sonar exposure. Collectively these studies show generally that blue whales may respond to controlled noise exposures in different ways, and that a suite of contextual factors influenced response probability. However, results from these kinds of studies are more challenging to apply directly within regulatory applications where more explicit individual information on response probability and severity are often required.

The above analyses of blue whale responses all involved methods assessing results across multiple individuals. These results demonstrate that some blue whales, which primarily use low frequency sound, may be sensitive to mid-frequency noise and that their responses appear to be influenced by various contextual factors. However, there is a further need to quantify individual responses (or lack of responses) of specified type and severity associated with known noise exposure conditions. Such data are directly useful in deriving exposure:response probabilistic functions for specific exposure variables commonly used in regulatory frameworks (e.g., received levels), as has been shown for Phase-I clinical trials in medicine and has been applied within other cetacean behavioral response studies (see: Miller et al., 2012; Southall et al., 2016). Individual case-by-case analyses also enable the evaluation of how other response covariates, such as source-individual range evaluated here, may also influence response probability (as in Harris et al., 2015). While this study includes individuals evaluated in a number of the studies above, by quantifying individual responses of blue whales to MFAS and PRN stimuli using whale-borne tags and CEEs we provide a completely novel analysis that is more explicitly applicable in predicting response probability in ways that are useful in regulatory decision-making. Further, comparing multiple methods that have been used in other studies provides an important evaluation across analytical methods for response analyses at the individual level to identify behavioral change-points for use in exposure:response functions.

II. METHODS

A. Study area and general field methods

This study was part of a long-term, multi-disciplinary research collaboration - the Southern California Behavioral Response Study (SOCAL-BRS). The CEEs presented here used several different experimental treatments with tagged blue whales during summer and autumn months (June-Oct) from 2010 to 2014 in coastal and offshore areas of the Southern California Bight. Within years, CEEs were conducted on different days (with two exceptions in 2010 where two CEEs were conducted within days at locations > 10 nm apart) in different geographical locations or spaced in time to the extent possible to reduce the occurrence of multiple exposures over short periods in the same area.

Detail on the SOCAL-BRS field methodology is provided in Southall et al. (2012; 2016) and is summarized here. Small (~6 m) rigid-hull inflatable boats (RHIBs) were used to locate, tag, and obtain positional and behavioral observational data for focal whales. A central research platform (*M/V Truth*; Truth Aquatics, Santa Barbara, CA) supported many research components, including the portable experimental sound source, passive acoustic listening systems, and visual observers on an elevated (7 m) observational platform directly above the ship's bridge. Visual observers supported RHIBs in locating focal whales and monitoring marine mammal exposures during CEEs to meet specified permit requirements. Individuals were identified visually and from photos in the field and in post hoc analyses to the extent possible using long-term photo identification records.

B. Quantifying individual blue whale behavior

Individual blue whale behavior was measured during phases defined as before, during, and after CEEs using a combination of high-resolution tag sensors and detailed focal follow procedures (see: Southall et al., 2012; Goldbogen et al., 2013). Tagging effort was concentrated on sub-adult or adult animals; no young calves (estimated by experienced field researchers as being less than six months of age) or mothers with young calves were tagged. Several types of motion sensing and acoustic tags were used. For the large majority of whales, DTAGs (version 2 and 3) (Johnson and Tyack 2003) were used. These tags included broadband hydrophones (<0.1 Hz – >100 kHz sensitivity) sampled at rates of 48-240 kHz depending on the tag type and configuration. Two whales in the first year of this experiment were tagged with B-Probes, sampled at rates of 20 kHz (see: Oleson et al., 2007). For each tag type, hydrophones were either calibrated directly or sensitivity was determined from calibrated tags of the same type. Acoustic records included environmental sounds, instances of calls produced by tagged and other whales (see: Goldbogen et al., 2014), known exposures to experimental stimuli, and other incidental anthropogenic noise including vessel noise and (in several instances) non-experimental military sonar of multiple types outside CEE periods. Tag-measured received levels (RLs) were quantified for both tag types using the same approach. The maximum RMS sound pressure level for each exposure stimulus within any 200 ms analysis window over the 1/3-octave band centered at 3.7 kHz, which contained the predominant sound energy of all exposure stimulus types (as in Tyack et al., 2011; Southall et al., 2012; DeRuiter et al., 2013; Goldbogen et al., 2013). Additionally, cumulative sound exposure levels (cSEL; in dB re: 1μPa^2-s) were measured as integrated sound energy across all received exposure stimuli (as in DeRuiter et al., 2013).

Fine-scale, three-dimensional movement data from individual diving, foraging, and other behavioral and kinematic parameters were obtained from pressure transducers and inertial measurement units at sampling rates from 5 to 250 Hz for DTAGs (Johnson and Tyack 2003) and 1 Hz for B-Probes (Goldbogen et al. 2006; Oleson et al., 2007). For the DTAGs with higher sample sensor resolution, the following tag-derived measurements were used for analyses: depth (m); absolute heading (degrees); heading variance (unitless); minimum specific acceleration (ms^{-2}); vertical and horizontal speed (ms^{-1}); lunges/dive; and feeding lunge rate (lunges h^{-1}). Heading variance was derived as relative variability between instantaneous absolute heading and median heading within each minute of tag data. Minimum specific acceleration (MSA) was derived from three-axis accelerometers as an integrated metric of overall acceleration (Simon et al. 2012). For the two B-probe deployments with lower sensor sample resolution, slightly different parameters were measured and used in analyses described below, including depth, fluking acceleration (ms^{-2}), and overall speed (ms^{-1}). For both tag types, the instantaneous velocity was determined by regressing the measured flow-noise from tags against the orientation-corrected changes in depth during stable ascending or descending portions of dives; this was calibrated for each individual tag deployment and tag orientation within the deployment (as in Cade et al., 2018). The instantaneous velocity was then multiplied by either the instantaneous pitch cosine (to obtain horizontal speed) or sine (for vertical speed) (Goldbogen et al., 2006). Feeding lunges were manually identified based on dive profiles, tri-axial body acceleration, and flow noise (as in Goldbogen et al., 2013). Given differential sensor sampling rates across tag types and sampling periods, all variables other than lunge rates were decimated to 1-Hz resolution. The minimum sampling rate across all tags (1 Hz) was sufficient to describe the most important biological relevant behaviors (feeding, diving).

Once animals were tagged, focal individual tracking commenced to obtain accurate spatio-temporal surfacing positions. Focal animal surface positions at known times were determined from either: known RHIB locations combined with range and bearing measurements to animals, measured from a precision laser range finder (Leica Vector, Viper II), known animal surface locations based on recent surface footprint locations, or in cases where direct measurements were not possible, visually estimated range and bearing from known RHIB locations to focal whales. Error in surface positions was estimated to be <10 m from directly measured locations and 10s to 100s of meters for visual estimates of range and bearing, depending on conditions and range from visual observers to whales. Focal whale positions were used to generate time-series maps of animal movement and relative (over-ground) speed estimates used in expert evaluation of potential response severity.

C. Synoptic environmental data

The overall vessel configuration and experimental paradigm were described in detail by Southall et al. (2012). However, subsequent to the original experimental design described therein was the inclusion of additional parameters related to the environmental contexts in which CEEs occurred.

Calibrated measurements of noise associated with SOCAL-BRS vessel operations were made under controlled, standardized conditions that were representative of typical field configurations. Remotely deployed drifting acoustic buoys supported passive acoustic recorders using both a primary surface float and an isolated smaller secondary float. Shock-reducing bungee cords were suspended from the secondary float, to which recorders were attached. Loggerhead DSG recorders (Loggerhead Instruments, Sarasota, FL, USA) were suspended to depths of ~30-m depending on the angle of the suspension line (small sea anchors were used to maintain a vertical orientation) and tension in the bungee. The DSG recording units were affixed with HTI-96 hydrophones (High Tech Inc., Long Beach, MS, USA) with a nominal sensitivity of -180 dB re 1 V/µPa and had a nominal 20-dB pre-amplifier gain; the recording unit had a resulting flat sensitivity of -160 dB re 1 V/µPa (+/-3 dB) between 16 Hz and 30 kHz. Recording buoys were deployed on three occasions in offshore locations (200-500m water depths) in areas near where CEEs were conducted. Recordings were obtained over three days in sea state 2-4 conditions; data presented here were obtained from the lowest possible sea state condition. Both RHIBs (*Ziphid* and *Physalus*) were instructed to pass by the surface float suspending recorders at a range of ~100m at speeds of 5 and 10 kts. This range was commonly the distance at which focal follows before, during, and after CEEs were conducted. The RHIBs traveled variable speeds during focal follows, depending on the behavior of the individual being followed, with 5 kts being a typical speed and 10 kts likely closer to a maximum speed. The central research vessel (*Truth*) was also instructed to pass recorders at ~100m range and speeds of 5-10 kts, which represented more of a worst-case scenario during CEEs (since the vessel was stationary and usually much further apart), but was more realistic in context of environmental prey mapping. Additionally, the *Truth* was instructed to position ~ 1 km from recorders and maneuver as if suspending the simulated MFAS sound source. These measurements provided received sound levels associated with the operation of the sound source vessel at typical ranges whales were during CEEs, in isolation from the experimental signals used in CEEs. For vessel passes, 1-min acoustic recordings centered on the time of the closest point approach (CPA) were selected for analysis. For each 1-min sample, one-third-octave band RMS levels (dB re 1 µPa) were then computed for each 1-s interval. Median values of all 60 samples were then calculated and are presented as representative noise levels that would be received by a whale at a relatively shallow depth (~30m) and in typical proximity during approaches from each vessel). For the stationary *Truth* maneuvering at ~1 km range from recorders, 2-min acoustic recordings during the confirmed time of maneuvering were used. Similarly, for each sample, one-third-octave band RMS levels (dB re 1 µPa) were computed for 1-s intervals. Median values of 120 samples were then calculated and are presented as representative noise levels that would be received by a whale at a relatively shallow depth (~30m) and in typical proximity during maneuvering of the *Truth* for sound source deployments during CEEs approaches. These values are then compared to comparable measurements of ambient noise made using the same and methods during the same day and similar conditions, with no experimental or other vessels operating within at least 3 km of recording buoys.

For some feeding whales during 2011-2014 CEEs, active acoustic methods were used to measure krill distribution and density in the proximity of feeding whales immediately before and after CEE sequences. The general approach in obtaining these measurements is described here; detailed methods for the collection and analyses of prey data are provided by Friedlaender et al. (2014, 2016) and Hazen, Friedlaender and Goldbogen (2015). Once a tag was deployed on a focal whale and as conditions allowed, a pre-exposure prey mapping survey was conducted at or near (typically within ~100 m) recent, known tagged whale surfacing positions. Across whales, this period lasted for 30-75 min prior to the onset of each full CEE sequence. This complete CEE sequence included three sequential 30 min phases (pre-exposure baseline, exposure, and post-exposure periods; see below), each of which occurred in the absence of active acoustic sampling (i.e., echosounders were not active during CEE sequences). Following the CEE sequence, a second 30-75 min active acoustic prey mapping survey was conducted. Given the clear importance of prey distribution in the behavior of feeding whales and in their responses during CEEs demonstrated by Friedlaender et al. (2016), we sought to evaluate the available prey distribution data in the context of potential responses even though contextual prey data were not available for all CEEs. Thus, we use prey data when available to provide additional context to the derived response likelihood that was conducted uniformly for all whales.

D. CEE methods

The experimental methods and specifications for the experimental sound source used in CEEs for this study are described in greater detail by Southall et al. (2012) and summarized within the context of other recent studies using CEEs to study behavioral responses of marine mammals to sonar by Southall et al. (2016). Essentially, a standard before-during-after (A-B-A) experimental design (with 30 min phases for up to a total of a 90 min full experimental sequence) was used to quantify potential changes in individual movement, diving, feeding, and other aspects of behavior where individual noise exposure was controlled and known.

Provided that numerous specific criteria were met regarding visibility, sea state, proximity to shore or other vessels, absence of very young calves, and other factors, the *Truth* was positioned at a range (typically 1000 m) estimated to result in maximum received RMS sound pressure level at the focal whale of 160 dB re 1µPa. In instances where multiple tagged whales were being monitored but were not in the same social group, a focal individual was selected in terms of positioning the sound source while a second tagged individual was followed by a second RHIB, but at some (typically greater) range that was less explicitly controlled. The experimental sound source was then deployed to a depth of 25 m and transmitted one of two signal types (MFAS: max 210 dB re 1µPa @ 1m or PRN: max 206 dB re 1µPa @ 1m) at 25 sec intervals during CEEs (see: Southall et al., 2012). Signals were ramped up from an initial source level of 160 dB re 1µPa @ 1m in 3 dB increments to the maximum source level for each respective signal type within the first ~7 min of exposure and were maintained at that level for the remainder of the CEE. Total exposure duration was a maximum of 30 min, but some exposure intervals were terminated early as a result of mitigation requirements (e.g., other animals swimming within 200 m of the active sound source) or because of equipment failure.

Following the completion of controlled noise exposure sequences, monitoring from archival tags and visual focal follow methods was maintained for at least 30 min. Early in this period, the experimental sound source was recovered, and the *Truth* was directed to maintain a comparable range (~1000 m) and speed relative to the focal whale (as done during the pre-exposure sequence). The RHIB maintained a comparable range and approach in the post-exposure as was done during the pre-exposure and exposure sequences. Complete CEE sequences thus consisted of constant monitoring using tags and visual follows of individuals from RHIBs during the consecutive 30 min pre-exposure, exposure, and post-exposure sequences. During these periods, the sound source vessel was mobile at a deliberately comparable range and relative orientation for the pre- and post-exposure but stationary (drifting) during the exposure period.

The primary research objective was to assess the potential responses of blue whales to military sonar. Consequently, and given the novelty of the study, a disproportionate number of CEEs, were conducted with MFAS stimuli. Following the first five exposure sequences during 2010 with MFAS, a 2:1 ratio of MFAS to PRN stimuli was used and tested in randomized order. While the primary experimental control was within the pre-during-post exposure experimental design, a smaller number of complete “control” sequences were conducted in which the full sequence was replicated and the sound source deployed but no noise stimuli were presented during the ‘exposure’ phase (Table 1).

In a single instance, a tagged blue whale was monitored while a CEE was conducted in coordination with an operational Navy ship (*USS Dewey*-DDG 105) using full scale MFAS (SQS-53C). Given the higher source level (235 dB re 1µPa @ 1m), *in situ* noise propagation modeling was conducted to position the vessel much further away from the individual in order to obtain the same desired maximum received level (~160 dB re 1µPa). A relative orientation was selected such that the ship was generally approaching the whale but was not directed precisely toward it and no course adjustments were made during transmissions. The ship transited a direct course at 8 kt and, given the inability to gradually increase the source level as was done with the experimental sonar, a slightly longer exposure period (60 min) with corresponding 60 min duration of pre-exposure and exposure phases was implemented.

Provided that tagged whales were being monitored according to specified criteria and conditions, CEEs were conducted irrespective of the animal's behavioral state at the time of exposure. To categorize each individual's behavioral state at the beginning of each CEE, the following *post hoc* criteria were used based on tag sensor data to define deep-feeding, shallow-feeding and non-feeding. The presence of a single foraging lunge during the baseline period was used to indicate a feeding state for the CEE. Any dive depth exceeding 50 m was used to distinguish deep from shallow diving.

Some CEEs were not fully completed, either due to tag failure or detachment, loss of visual contact with individuals for long periods, or premature termination of noise exposure resulting from required termination protocols or equipment failure. Because of the difficulty in obtaining large sample sizes for such experiments under field conditions, incomplete sequences were retained within partial analyses when possible. Where individuals were successfully monitored with tags and visual observations through the pre-exposure and at least half (15 min) of the experimental period, the CEE was included. Behavioral response analyses were conducted, although without the ability to evaluate potential recovery from any responses during post-exposure periods. This is an additional benefit of individual-based, time-series analyses over a synthetic analytical approach.

E. Behavioral response analyses

Individual blue whale behavior and potential responses during noise exposure periods were evaluated in parallel using two different analytical approaches: a structured expert evaluation and a quantitative statistical analysis. Methods for each are discussed below and results are presented within each analytical method by individual and evaluated together based on CEE stimulus type and animal behavioral state at the start of CEEs.

i. Expert scoring analyses

A structured evaluation of selected, standardized data streams using method derived by Miller et al. (2012) based on the Southall et al. (2007) response severity scaling developed by was conducted by two independent groups of subject matter experts, each containing three of the co-authors (1: AF, AS, JG; 2: JC, AA, GS). Each group was provided synoptic time-series behavioral information in the form of annotated maps of individual spatial movement (from RHIB-based focal follows) and selected kinematic and behavioral parameters in time-series plots (extracted or derived from tag records). For DTAGs (40 of 42 individuals), these included: depth (m); feeding rate (lunges $dive^{-1}$); MSA (ms^{-2}); absolute heading (degrees), and horizontal speed (ms^{-1}). For the two BProbe deployments, these included depth (m), fluking acceleration (ms^{-2}), and overall speed (ms^{-1}). As in Miller et al. (2012), many of the scorers were involved in the original fieldwork and thus may have had some recollection of events during CEEs (although some occurred over four years prior to expert scoring). In order to minimize any biases resulting from experience, scorers in this study were blind to the individual whale ID, date and location of CEEs, exposure treatment, or precise timing of received levels of exposure signals and CEEs were presented to groups in randomized order in terms of the date that the experiment was conducted. Experimental phases (pre-, during-, post-exposure) for each CEE were identified in all data plots provided to each scoring group. This allowed scorers to evaluate behavior in pre-exposure baseline conditions, identify potential behavioral changes during exposure at specified times, and to assess whether any identified behavioral changes persisted throughout and/or after noise exposure. The three members of each group collectively evaluated these data plots and annotated maps and time-series data plots for each CEE. Maps showed the position of the experimental sound source at the start and end of the CEE, every surface location collected by RHIBs during individual focal follows identified in each CEE phase (with times shown for the first position in each phase), and a 1000 m radius around the source at the onset of exposure for scale.

Scorers were instructed to evaluate the annotated maps and data plots for each CEE and to identify any behavioral changes to the nearest minute that occurred based on the descriptions specified in the severity scale. Criteria for temporal descriptors were as follows: brief or minor changes were identified as those returning to baseline conditions during exposure; moderate duration changes were identified as those not returning to baseline conditions until into the post-exposure period; extended duration changes were those not observed to return to baseline within the post-exposure period. If multiple changes were identified, all were reported based on visual inspection of plots. The two groups independently evaluated each CEE collectively and came to a consensus agreement about any identified behavioral changes, the time at which they occurred, and a confidence level (low, moderate, high) as to the overall severity score(s) for each CEE. Where no behavioral responses were identified, a severity score of 0 was assigned. Where multiple responses were identified, all were reported, but the most severe (highest score) was used as the resulting overall score for that CEE. Neither Southall et al. (2007) nor Miller et al. (2012) identified an increase in feeding as an adverse behavioral change. Because this was not included within the severity scale, when it occurred it was not systematically reported and scored by expert scoring groups here. It was noted on multiple occasions as a change but was not scored as an adverse reaction.

After each group independently completed their evaluation of all CEEs, both groups met to compare results. An adjudicator (BLS) was selected to mediate the combined group discussion and served to break any irreconcilable disagreements that occurred about severity scores between groups. A consensus behavioral response severity score (0 for no response; 9 for most severe response), a confidence score (low, med, high), and specified exposure times for any changes, were identified for all MFAS, PRN, and control (no noise) sequences. If a behavioral response was identified, the time of the response was used to derive exposure RLs (max RMS and cSEL to that point within the CEE).

Exposure-response probability functions were then generated using recurrent event survival analysis to assess time-to-event changes using marginal stratified Cox proportional hazards models fitted to the severity score data (see: Harris et al., 2015 for full details of model application to severity score data). These models combine the results from individual CEEs to estimate the likelihood of response as a function of exposure received level (in cSEL) and behavioral or contextual covariates. Models were fitted to broad categories of response severity levels (i.e., low, moderate, high) to ensure sufficient data to support the dose-response functions. The resulting hazard models provide a relationship between exposure level and the probability of response at different severity levels, while accounting for selected contextual variables. Similar analyses have been conducted for pilot whales, killer whales and sperm whales (Miller et al., 2012; Harris et al., 2015), as well as humpback whales (Sivle et al. 2015).

Given data limitations for shallow and non-feeding behavioral states, the Cox proportional hazards models were only fitted to data from animals that were deep feeding in the pre-exposure period. For these cases, the first occurrence of each response level (severity scores 1-3, 4-6, 7-9) was determined based on consensus expert scored results for each CEE for inclusion in the models. For CEEs with a severity score of 0 (no response), the cSEL for the entire exposure sequence was used and the data were labeled as right-censored, meaning that no response was detected up to this exposure level. We fitted models to data from all CEEs associated with deep feeding animals and included source-animal range (m) at the start of the exposure phase and signal type (MFAS or PRN) as covariates. Observations were assumed to be correlated within individuals but independent between individuals. The standard errors of the model estimates were corrected for the correlations within individuals using a grouped jack-knife procedure (Therneau and Grambsch 2000). All possible model combinations from the null model through to two-way interaction terms were fitted and AIC-based model selection was used. For the selected model, the proportional hazards assumption was verified (Kleinbaum and Klein 2005; Harris et al., 2015). Analyses were conducted in R version 3.0.2 (R Core Team, 2013) and exposure-response functions were generated as survival curves from the fitted models using the survfit function package (Therneau 2014).

b. Mahalanobis distance (MD) statistical analyses

A Mahalanobis distance (MD) method (Mahalanobis, 1936; see: DeRuiter et al., 2013) was also used to statistically test for change-points in whale behavior. This approach involves the calculation of an integrated statistical distance-based metric that summarizes synoptic dive parameters from tag data and quantifies how they differ over time from those present within a specified baseline period (e.g., pre-exposure period). The MD metric is a scale-invariant integrated 'difference' from baseline behavioral parameters calculated in multi-dimensional space and accounting for correlations between dimensions. It is calculated within a sliding temporal window across all dive parameters to identify the specific time (if any) at which overall behavior changed. A window duration of 5 min (a conservative average dive duration for blue whales across all behavioral states) was selected with an MD value calculated every 25 seconds (corresponding to the interval between the onset of individual noise transmissions during CEEs). The MD calculations require a variance/covariance matrix to quantify statistical relationships among all variables. We calculated this matrix for each whale using the full dataset for the entire deployment, excluding an initial 15-min. period estimated (based on nominal blue whale diving behavior) to account for any tagging effects (based on Miller et al., 2009). The inclusion of the full dataset, including and following CEE periods, was deemed necessary to provide sufficient samples to accurately estimate matrix parameter values. It was also considered a conservative choice, in that if behavioral changes during or following exposure were such that the variance-covariance structure was altered, the MD analyses would be less likely to detect it when using the full dataset than if only pre-exposure data had been used.

The following behavioral parameters (all quantified from individual animal-borne tags) were used as input variables in calculating MDs. For DTAGs this included: circular variance of heading (25 sec window); MSA (ms^{-2}); vertical speed (ms^{-1}); horizontal speed (ms^{-1}); feeding lunge rate (lunges h^{-1}, 15 min window), all at 1 Hz resolution. For the two Bprobe deployments, this included: overall speed (ms^{-1}); and feeding lunge rate (lunges h^{-1}) at 1 Hz resolution. Dive data from the 30-min pre-exposure period (where other contextual factors including experimental vessel presence were similar to those during exposure) were used as comparison baseline data; this period also began at least 15 min post-tagging. When a tagged whale was near the surface, all data points that were collected shallower than 10 m were replaced with median parameter values from the baseline period to result in MD values near zero. This was to account for artifacts introduced by noise in some input data streams, most notably accelerometer-based metrics. This effectively pulls MD values toward 0 as the proportion of data points obtained at shallow depths in a time-window increases. The MD was then computed between (1) average behavioral data parameters for the baseline period and 2) average data values within the 5 min sliding comparison window.

Exposure and post-exposure periods were then evaluated to determine whether an individual behavioral change occurred, when it began, and when it ended. MD values exceeding the maximum value observed during the pre-exposure period were identified as behavioral changes. For consistency with the expert scoring severity assessment, detected changes associated with the onset of or increase in foraging were not considered responses that would have any potential negative effects for individuals. Therefore, they were not included in the expert severity scoring options and were not reported as detected changes.

III. RESULTS

A. CEE Results

A total of 48 CEE sequences were conducted for individual whales involving MFAS, PRN, or no noise 'control' exposures in (primarily) coastal and offshore areas spanning the southern California Bight (Fig. 1). Data from six sequences in which tags detached prematurely or CEE sequences were terminated before 15-min of exposure were not included in this analysis as they failed to meet specified experimental criteria; the remaining 42 sequences met these criteria and were analyzed. These occurred within 33 discrete CEEs, as nine of these sequences involved two concurrently tagged and followed animals. In seven of these instances, simultaneously tagged whales were separated from one another and were followed by separate boats. In two cases, simultaneously tagged individuals occurred within close proximity and were being tracked within the same focal follow, although one of these the animals was later determined to be in different behavioral states during exposure. Four individual whales were later revealed through photo identification to have been exposed in two separate CEEs within the same year. In each scenario, CEEs were spaced by several days or weeks. Furthermore, in each case animals received different treatment types and were in different behavioral states for subsequent exposures. This likely reduced, but did not eliminate, the potential that behavioral responses during the second CEEs may have been influenced to some degree by exposure to the initial ones.

The 42 discrete, randomized CEE sequences evaluated here were conducted during 2010-2014 field efforts within different exposure treatments and behavioral state contexts. The resulting distribution of CEEs conducted for individuals within these three different behavioral states for each treatment type are summarized in Table 1. Representative examples of different types of behavioral response results for three individual whales are provided (Fig. 2).

The results of CEE #2011-01 on 29 July 2011 with individual bw11_210b are shown in time-annotated maps and MD data plots with received cSEL (in dB re 1μPa^2-s) in the top panels (Fig. 2 a,b). This was a deep-feeding blue whale exposed to MFAS at a source-whale horizontal range (at the start of the exposure) of 1.2 km. Clear changes in behavior were detected with both MD and expert scoring methods (high confidence) at virtually the same time (1528-1529 PDT), corresponding to a received cSEL of 119 dB re 1μPa^2-s. Changes identified by adjudicated expert scoring included horizontal avoidance of sound source (severity score 7) and moderate cessation of feeding (6) (see Table S1 for expert scoring details). The results of CEE #2011-06 on 6 August 2011 with individual bw11_218b are shown in the middle panels (Fig. 2 c,d). This was a deep-feeding blue whale exposed to PRN at a source-whale range (at the start of the exposure) of 5.6 km. No changes in behavior were detected with either MD or expert scoring methods (high confidence) despite a relatively high received cSEL of 168 dB re 1μPa^2-s (see Table S1 for expert scoring details). The results of CEE #2013-06 on 26 July 2013 with individual bw13_207a are given in the bottom panels (Fig 2 e,f). This was a shallow-feeding blue whale within a control sequence conducted at a source-whale range of 0.5 km. No changes in behavior were detected with expert scoring methods (moderate confidence), although the presence of increased feeding was noted (see Table S1 for expert scoring details). The increase in feeding rate resulted in a gradual increase in the MD metric relative to the pre-exposure baseline condition and was thus detected as a change. As in several other instances where whales initiated or increased feeding during CEEs, the MD detected change was noted, but was not considered a conflicting result to the expert scoring evaluation because an increase in feeding was not defined as an adverse behavioral response (Southall et al., 2007; Miller et al., 2012).

Expert scoring and MD results are presented for each treatment type and behavioral state category for each individual blue whale (Table 2). Received exposure levels for each whale either at identified change points or (where none were detected) maximum values for CEE sequences are also provided (Table 2). For CEEs with identified responses cSEL values at identified change points ranged from 97 to 155 dB re 1μPa^2-s. Maximum cSEL values for CEEs where no change was identified ranged from 134 to 171 dB re 1μPa^2-s. Source-whale range varied from 0.4 to 7.7 km for the simulated MFAS and 19.5 km for the single operational vessel MFAS signal, with a median range of 1.2 km. There was no significant correlation within experimental sound types (MFAS, PRN) across CEEs between received level and source-whale range.

i. Deep-feeding whales

The largest number of individual CEE sequences analyzed (n=29) occurred for blue whales engaged in deep-feeding during pre-exposure periods. Whales were most likely to respond during MFAS CEE sequences, with a similar overall proportion of individuals identified as changing behavior during exposure by both expert scoring (8 of 13) and MD (9 of 13) methods. A lower proportion of deep-feeding whales responded when exposed to PRN (4 of 11 in expert scoring analysis, 5 of 11 for MD) and almost no responses were detected in deep-feeding control sequences (0 of 5 for expert scoring; 1 of 5 for MD).

For a subset of deep-feeding whales (n=21), prey distribution and density were measured before and after CEE sequences to provide an environmental context for interpreting responses in this behavioral state. Given the knowledge of the importance of this contextual relationship, we include three examples of whale behavior and contextual prey data to illustrate how these measurements provide additional insight into changes in whale behavior and the interpretation of potential response (Fig. 3).

For bw11_210b on 29 July 2011 (Fig. 3a; Fig S3), prey patch depth and density remained similar both before and shortly following the CEE (#2011-01) in the area where the whale was feeding. Both expert scoring groups identified very similar behavioral changes with high confidence at approximately the same time as one another and as the MD analysis (see Table S1 for expert scoring details), which identified a clear change relative to not only the pre-exposure condition, but the entire behavioral record for this individual (including pre-CEE prey sampling periods). Given the similarity in the prey environment before and at least immediately after the CEE, these identified changes (avoidance and cessation of feeding) are unlikely the result of changes in the prey environment (from the exposure or otherwise). However, subsequent changes in the overall prey environment (more schools identified at various depths) and/or changes in the local prey environment based on the whale's geographic location may have also influenced whale behavior, particularly well after the CEE.

For bw11_218b on 6 August 2011 (Fig. 3b; Fig S4), prey patches after the CEE (#2011-06) were shallower than those measured before the CEE sequence. This whale appeared to progressively decrease its feeding depth and continue to feed during the CEE as it moved into an area with shallower patches. This gradual decrease in whale diving depth was not identified by either expert-scoring group as a behavioral response during the CEE (Table S1). A behavioral change point was identified within the MD analysis (See Fig. S4 where the MD trace crosses the dashed line representing the pre-exposure baseline value used as the response threshold), although this was a small increase above the pre-exposure baseline period and it was of smaller magnitude than the MD spike in this metric identified just after the pre-CEE prey sampling period.

For bw13_207a on 26 July 2013 (Fig. 3c; Fig S5), prey patches measured around the CEE (#2013-06) in the area where the whale was feeding were deeper and less dense following the CEE sequence than before exposure. The animal maintained a similar feeding depth before and during the exposure sequence but increased its feeding rate and switched to deeper feeding after the CEE, which also continued during the post-exposure prey sampling period. Neither expert scoring group identified any behavioral change in this CEE, but there was a discernable change detected using the MD method, associated with an increase in foraging during the exposure phase relative to the defined baseline (pre-exposure) period (see Table S1 for expert scoring details). These MD values were of similar magnitude to those measured during both prey sampling periods (before and after the full CEE sequence).

Cox proportional hazards models were fitted separately to responses of severity scores between 4-6 and 7-9; responses with severity scores of 1-3 were insufficient to apply this process. The Cox proportional hazards model selected by AIC for severity score 4-6 retained only source-whale range as a covariate (Δ AIC=1.34), although its effect was not significant (p=0.316). The selected model met the proportional hazards assumption (global p-value from Chi-square test = 0.079). The model selected by AIC for severity score 7-9 was the null model (Δ AIC=1.03), with the model including source-whale range being the second best model according to AIC. Given the interest in understanding the role of source-whale range in the probability of responding, model results from the selected model for severity scores between 4-6 and the second-best model for severity scores between 7-9 were used to produce predicted exposure-response probability functions in terms of received exposure level for the two different response severity levels (moderate severity: 4-6; high severity: 7-9). In order to illustrate the relationship with source-animal range, response probability functions were calculated for the ranges over which most CEEs were conducted (1-5km) (Fig. 4). These prediction plots suggest that the probability of a moderate response (severity 4-6) as a function of RL decreases rapidly as range increases, but the wide confidence intervals indicate substantial uncertainty in this relationship. The relationship is much less pronounced for high severity responses (severity 7-9) hence the selection of the null model.

ii. Shallow-feeding and non-feeding whales

The second largest number of individual CEE sequences analyzed (n=8) occurred for blue whales engaged in shallow-feeding during pre-exposure periods. No whales (0 of 7) were determined to change behavior during MFAS exposure by either expert scoring or MD methods. No PRN sequences were conducted for shallow-feeding whales. No responses were detected by either analytical method during the single shallow-feeding control sequence.

The fewest number of individual CEE sequences analyzed (n=5) occurred for non-feeding blue whales, although most of these individuals were determined to have an adverse behavioral response during CEEs across both methods. For MFAS CEE sequences, expert scoring determined such a response in one of two whales while MD analyses detected adverse responses for both individuals. For PRN CEEs, expert scoring determined an adverse behavioral response in one of three non-feeding whales whereas all three individuals were identified to have such a response using MD methods. No control sequences were conducted for non-feeding whales.

B. Vessel noise characterization

Median values of vessel noise were calculated for CPAs for all vessels during each condition. These values were compared for each condition for RHIBs *Ziphid* and *Physalus* to comparable measurements of ambient noise made using the same recorders and methods during the same day and similar conditions, with these vessels operating at much further ranges from recording buoys (Fig S1). Ambient noise measurements were also compared for each passage condition for the *M/V Truth* to comparable measurements of ambient noise made using the same recorders and methods during the same day and similar conditions, with this vessel operating at much further ranges from recording buoys (Fig S2a, b). For the stationary *Truth* maneuvering at ~1 km range from recorders, median noise values were calculated relative to ambient noise during the same day and similar conditions (Fig S2c). Both RHIBs and the *Truth* were clearly detectable over ambient noise for both speeds at these close ranges, with different relative spectral distribution of noise energy at different speeds for each vessel. Based on the associated noise levels and frequencies and typical ambient noise during non-vessel periods, their operation is likely audible to subjects over ranges typical during CEEs, particularly the RHIBs at their typical operating speeds and ranges from animals. However, as a part of the experimental design during baseline, exposure, and post-exposure sequences, these represent relatively continuous levels of additional noise exposure. During sound source deployment, the *Truth* conducted small maneuvers to remain stationary. The measurements of ambient noise during this period demonstrated that these maneuvers and the presence of the vessel was not discriminable over noise measured using the same recording system in the absence of the *Truth.* That is, while vessels were likely audible during their operation, particularly during pre- and post-exposure periods when the Truth was following focal animals, during exposure periods from the sound source vessel received by experimental subjects was predominately or exclusively the result of experimental exposures.

IV. DISCUSSION

This study generated the largest sample size (n=42) for any experimental behavioral response study involving sonar conducted to date for any marine mammal species (Southall et al., 2016). While the number of individual CEEs conducted in some behavioral states and treatments were limited, dozens of controlled individual experiments were conducted using high-resolution movement and acoustic sensors for individuals in well-defined exposure contexts. These results provide direct and robust means of evaluating how an endangered species responds to noise exposure, including simulated and actual military MFAS signals that have been associated with lethal responses in other species. The analytical approach provides a direct means of quantifying individual behavior and behavioral responses within known noise exposure conditions in such a way that probabilistic response functions may be generated in light of important contextual variables. Such data provide an empirical basis for modeling efforts to evaluate potential consequences of disturbance at broader population scales (King et al., 2015; McHuron et al., 2018; Pirotta et al., 2018).

Blue whales responded to noise in some but not all CEE sequences (19 of 37 for MD analysis; 14 of 37 for expert scoring) and in almost no control (no-noise) sequence (1 of 6 for MD analysis; 0 of 6 for expert scoring). Treatment types had variable sample sizes, but responses were generally equally likely to occur for MFAS and PRN exposures. Other than a single instance detected only with the MD method, none occurred during control (no noise) sequences. Nine CEEs involved exposure of multiple individuals, although nearly all of these included animals in separate groups. A small number of CEEs involved paired individuals or subsequent exposures to the same individuals and in two instances in the first year of the study animals could have been remotely exposed to an earlier CEE prior to being the focal animal in a subsequent CEE later in the day. While these could call into question the treatment of all individuals as independent samples, they were treated as such here (rather than excluding individuals) given the small number of instances relative to the overall sample size. Further, we took into consideration the fact that in all but one instance these CEEs all involved differences in individual behavioral state and/or treatment type.

Responses generally included short-term changes in diving behavior, small-scale (few km) horizontal avoidance of sound source location, and/or cessation of feeding activity. Recovery to typical pre-exposure behavior in most CEEs typically occurred within the post-exposure phase. However, the short-term and relatively rapid nature of recovery should be considered within the context of acknowledged differences between the MFAS from an experimental source and operational MFAS. The experimental MFAS is stationary, includes a ramp-up escalation of the source level, and the overall duration is relatively short (tens of minutes). Operational MFAS training involves much louder and constant levels and can occur over many hours or even days in the case of multi-ship operations (see: Moretti et al., 2014). It can also occur at any hour of the day and throughout the year, whereas CEEs here were only conducted during daylight hours in the summer and autumn.

Two different analytical approaches were applied to evaluate behavioral changes from baseline conditions within individuals using high-resolution, time-series kinematic and acoustic data. This approach included both quantitative statistical change-point methods and structured expert scoring assessment of deviations from baseline conditions during exposure by subject matter experts. The MD method is inherently objective in that it simply identifies changes in a suite of variables from baseline (pre-exposure) conditions and is thus equally likely to detect a behavioral change associated with a presumably positive outcome (e.g., an increase in foraging behavior) as a presumably negative outcome (cessation of feeding). Further, the selection of a response "threshold" for MD strongly affects the probability of statistically detecting a behavioral response. Here a fairly low MD value was selected as a change-point threshold, namely a MD value within the exposure period exceeding that measured during the pre-exposure period. This results in a higher likelihood of identifying a behavioral response than if an alternate threshold were selected (e.g., two standard deviations exceeding the pre-exposure maximum) or if MD values during exposure exceeded the pre-exposure maximum value across the entire tag record. However, the intent here was to identify a discernable change in behavior during an exposure period with a similar context as pre-exposure conditions (e.g., local environmental variables, proximity of vessels) rather than aiming to identify a change that was more unusual than any other change measured for that or any other blue whale. Not surprisingly, the MD method was more likely to detect a change than expert scoring, both in controls and exposures. However, once detected changes associated with the onset of feeding (presumably not an adverse behavioral change) were discounted, results were quite similar across individuals. Some differences were still observed, but for 32 of 42 CEEs (76%), the methods agreed as to whether an adverse behavioral change occurred (where changes associated with the onset of feeding were excluded). Further, detected changes tended to occur at similar exposure times and associated received levels. Expert scoring methods were consequently consistent with the MD method in identifying behavioral changes, but this approach also has the advantage of being descriptive and identifying changes associated with various types of behavior (movement, feeding), including variability in response severity and the level of confidence in discerning response both within and between groups. While both methods have advantages and limitations, the general agreement here was encouraging and having used both methods provides more comprehensive insight into changes during experimental exposures. Future studies should consider integrating objective statistical change-point analyses (e.g., MD results) within expert evaluation of potential responses.

These findings demonstrate the kinds of context-specific differences in behavioral response identified by Ellison et al. (2012). Along these lines, they also complement and expand upon the findings of Goldbogen et al. (2013) and DeRuiter et al. (2017) regarding the importance of behavioral state in terms of response probability for blue whales, specifically the increased likelihood of response in deep-feeding animals. This study provides a different perspective on this behavioral state dependency in evaluating individual response type and severity for known exposure conditions for a relatively large sample size. Given these observations, we note the contextual differences between the simulated MFAS and some kinds of operational MFAS sources such as the SQS-53C sonar used in one CEE here; there are greater contextual similarities between the experimental source and other common operational military MFAS sources such as helicopter-dipping sonars. The experimental MFAS has proven useful in demonstrating previously unknown aspects of behavior, response, and context-dependency in these species, but, as we’ve shown, differences in exposure parameters can influence response probability. Additional research, some of which has been conducted and some of which is underway, is needed to further evaluate the importance of contextual differences in sound source type (e.g., source level, movement, spectral features) and proximity. This approach with individual animals where exposure range was known allowed for a quantification of behavioral response probability as a function of proximity to the sound source (Fig. 4) for the ranges tested. For deep feeding animals, whales had a higher response probability when located closer to the sound source for comparable RLs, although there is considerable uncertainty within the relationships and insufficient data to test this relationship for other behavioral states. Given the available data at this point, a simple relationship between source range, received level, and response probability across all whales does not appear to exist. Further evaluation of the potential range-dependence identified within this study using a dedicated experimental design to test and further resolve these seemingly important range-received level relationships is needed before firm conclusions can be drawn. Specifically, additional studies should explicitly evaluate different dimensions of the received level-range space, including potential changes during near but quieter exposure conditions.

Whale dive depth has been closely linked to changes in prey patch depth, thus prey can both mediate the response to sonar playbacks when prey are dense and can confound potential responses when prey distributions are not known. While a direct quantitative comparison is not possible for all individuals, given the absence of before and after prey data in some cases, our results were consistent with Friedlaender et al. (2016) in suggesting that the behavior of feeding blue whales is broadly influenced by features of the prey environment in ways that likely mediate responses to CEEs. Specifically, two of the three instances where the MD detected CEE responses were potentially a result of changes in prey while expert scoring classified 0 of the 3 as a CEE response (see supplemental materials for additional details). This highlights a potential strength in expert scoring in identifying specific aspects of a response in the absence of known important contextual variables. Changes in prey patch depth have been shown to result in commensurate changes in whale dive depth, and for some individuals, the likelihood of a behavioral response to navy sonar during a playback is reduced with increased prey density while foraging.

Many regulatory efforts to evaluate the effects of noise on marine mammals have primarily or exclusively used received noise exposure level as a predictor of response probability and have sought to develop more robust predictive associations. As illustrated by Ellison et al. (2012), a host of contextual factors can influence behavioral responses to noise. Several key contextual influences were identified here (and see: Goldbogen et al., 2013; Friedlaender et al., 2016; DeRuiter et al., 2017) that have strong effects on whether and how endangered blue whales respond when exposed to military MFAS signals or PRN of similar frequency and duration. Responses were mediated by a complex interaction of the animal's behavioral state at the time of exposure, features of the environment, and the relative proximity of sound sources. Without identifying behavioral state using objective, quantitative metrics (e.g., dive depth, presence of foraging lunges) and considering this as a relevant contextual variable, it would have been much more difficult to unravel the complexity of these relationships across studies. Identifying this, within certain contexts, indicates that an increase in received levels are in fact associated with an increase in response probability. While this complexity is not yet fully understood, relating response probability, exposure level, and behavioral state dependency will enable a more insightful and informed understanding of exposure-response relationships. This does not mean that each behavioral state and/or prey contextual condition must be informed by distinct and empirical exposure-response risk functions for management applications. Rather, integrated risk functions within behavioral states (e.g., foraging, traveling) and a small subset of contextual covariates (e.g., range) might be informed by targeted experimental studies in some species where relatively large sample sizes may be obtained (see Southall et al., 2016; Southall, 2017).

These results provide further evidence and increased resolution on how baleen whales respond to noise exposure. They also provide much-needed direct measurements of behavioral responses in an endangered species commonly exposed to MFAS within important habitat areas off California. As has been noted in other studies (see: Southall et al., 2016; Southall, 2017), results from locations where sonar exposure is common is likely much different from the behavioral responses of animals from areas where sonar exposure is uncommon or absent. Although blue whales are likely low-frequency specialists, they can and do respond to sounds presented to them with primary energy in the 3-4 kHz range associated with many MFAS systems found in commercial, naval, and recreational platforms. Whales that do respond appear to recover to typical behavioral patterns relatively quickly based on the results from these CEEs, and their probability of response should be considered given the contextual dependencies described in this study. With increased energetic demands and needs for high density prey, even the cessation of feeding for a short time could have consequences for the fitness of these large animals (see: Goldbogen et al., 2013). If they are chronic, they could manifest as population-level effects. Future experimental studies and targeted monitoring informed by these results should focus on the energetic and, in turn, biological consequences of behavioral responses across different behavioral states.

ACKNOWLEDGMENTS

This five-year study represented a major portion of the overall SOCAL-BRS project and could not have been conducted without the tireless support of many dedicated field personnel. We sincerely appreciate the Truth Aquatics team for their sustained involvement and support. We would also like to specifically acknowledge the contributions of: Kristin Southall, Peter Tyack, Jay Barlow, Shannon Rankin, Fleur Visser, Annie Douglas, Erin Falcone, Katy Laveck, Jeff Foster, Todd Pusser, Glenn Gailey, Doug Nowacek, Louise Burt, John Hildebrand, Ron Morrissey, Greg Juselis, Jolie Harrison, Tami Adams, Sarah Wilkin, Ned Cyr, Frank Stone, Bob Gisiner, and Mike Weise. All research activities for this study were authorized and conducted under U.S. National Marine Fisheries Service permit #14534; Channel Islands National Marine Sanctuary permit #2010-004; U.S. Department of Defense Bureau of Medicine and Surgery (BUMED) authorization; a federal consistency determination by the California Coastal Commission; and numerous institutional animal care and use committee (IACUC) authorizations. We also thank two anonymous reviewers who greatly improved the quality of the manuscript.

COMPETING INTERESTS

No competing interests declared.

FUNDING

Primary funding for the SOCAL-BRS project was initially provided by the U.S. Navy's Chief of Naval Operations Environmental Readiness Division and subsequently by the U.S. Navy's Living Marine Resources Program. Additional support for environmental sampling and logistics was also provided by the Office of Naval Research, Marine Mammal Program. Additional support to enable S. Guan's participation was provided by the U.S. National Oceanic and Atmospheric Administration.

LIST OF SYMBOLS AND ABBREVIATIONS

CEE	controlled exposure experiment
cSEL	cumulative sound exposure level
MD	Mahalanobis distance
MFAS	mid-frequency active sonar
MSA	minimum-specific body acceleration
PRN	pseudo-random noise
RHIB	rigid-hull inflatable boat
RL	received level

SOCAL-BRS Southern California Behavioral Response Study

REFERENCES

Cade, D. E., Barr, K. R., Calambokidis, J., Friedlaender, A. S., & Goldbogen, J. A. (2018). Determining forward speed from accelerometer jiggle in aquatic environments. ***J. Exp. Biol.* 221, jeb170449.**

Cooke, J.G. 2018. *Balaenoptera musculus*. *The IUCN Red List of Threatened Species* 2018: e.T2477A50226195.

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. and Thomas, L. (2013). First direct measurements of behavioural responses by cuvier 's beaked whales to mid-frequency active sonar. *Bio Lett.* ***9***, 20130223.

DeRuiter, S. L., Langrock, R., Skirbutas, T., Goldbogen, J. A., Calambokidis J., Friedlaender, A.S., and Southall, B. L. (2017). A multivariate mixed hidden markov model for blue whale behaviour and responses to sound exposure. *Ann. Appl. Stat.* **11,** 362–92.

Ellison, W. T., Southall, B. L., Clark, C. W., and Frankel, A. S. (2012) A new context-based approach to assess marine mammal behavioral responses to anthropogenic sound. *Conserv. Biol.* **26,** 21– 28.

Falcone, E.A., Schorr, G.S., Watwood, S.L., DeRuiter, S.L., Zerbini, A.N., Andrews, R.D., Morrissey, R.P. and Moretti, D.J. (2017). Diving behaviour of Cuvier's beaked whales exposed to two types of military sonar. *Royal Society open science*, ***4*,** p.170629.

Filadelfo, R., Mintz, J., Michlovich, E., D'Amico, A., Tyack, P. L., and Ketten, D. R. (2009). Correlating military sonar use with beaked whale mass strandings: What do the historical data show? *Aquat. Mamm.* **35**, 435–444.

Friedlaender, A. S., Goldbogen, J. A., Nowacek, D. P., Read, A. J., Johnston, D., and Gales, N. (2014). Feeding rates and under-ice foraging strategies of the smallest lunge filter feeder, the Antarctic minke whale (*Balaenoptera bonaerensis*). J. Exp. Biol. **217**, 2851–2854.

Friedlaender, A. S., Hazen, E. L., Goldbogen, J. A., Stimpert, A. K., Calambokidis, J., and Southall, B. L. (2016). Prey-mediated behavioral responses of feeding blue whales in controlled sound exposure experiments. *Ecol. Appl.* ***26***, 1075-1085.

Fristrup, K. M., Hatch, L. T., and Clark, C. W. (2003). Variation in humpback whale (*Megaptera novaeangliae*) song length in relation to low-frequency sound broadcasts. *J. Acoust. Soc. Am.* **113**, 3411-3424.

Goldbogen, J. A., Calambokidis, J., Shadwick, R. E., Oleson, E. M., McDonald, M. A., & Hildebrand, J. A. (2006). Kinematics of foraging dives and lunge-feeding in fin whales. *J. Exp. Biol.* **209**, 1231–44.

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. and Kyburg, C. 2013. Blue whales respond to simulated mid-frequency military sonar. *Proc. R. Soc. B*, ***280*****(1765)**, p.20130657**.**

Goldbogen, J. A., Stimpert, A. K., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Schorr, G. S., Moretti, D. J., Tyack, P. L., and Southall, B. L. (2014). Using accelerometers to determine the calling behavior of tagged baleen whales. *J. Exp. Biol.* **14,** 2449–2455.

Harris, C. M., Sadykova, D., DeRuiter, S. L., Tyack, P. L., Miller, P. J. O., Kvadsheim, P. H., Lam, F. P. A., and Thomas, L. (2015). Dose response severity functions for acoustic disturbance in cetaceans using recurrent event survival analysis. *Ecosphere* **6**, 1–14.

Hatch, L. T., Wahle, C. M., Gedamke, J., Harrison, J., Laws, B., Moore, S. E., Stadler, J. H., and Van Parijs, S. E. (2016). Can you hear me here? managing acoustic habitat in us waters. *Endanger. Species Res.* **30**, 171–186.

Hazen, E. L., Friedlaender, A. S., and Goldbogen, J. A. (2015). Blue whales

(*Balaenoptera musculus)* optimize foraging efficiency by balancing oxygen use and energy gain as a function of prey density. *Sci. Adv.* **1**, e1500469–e1500469.

Henderson, E. E., Smith, M. H., Gassmann, M., Wiggins, S. M., Douglas, A. B., and Hildebrand, J. A. (2014). Delphinid behavioral responses to incidental mid-frequency active sonar. *J. Acoust. Soc. Am.* **136**, 2003–2014.

Isojunno, S., Cure, C., Kvadsheim, P. H., Lam, F. P. A, Tyack, P. L., Wensveen, P. J., and Miller, P. J. O. (2016). Sperm whales reduce foraging effort during exposure to 1–2 khz sonar and killer whale sounds. *Ecol. Appl.* **26**, 77–93.

Johnson, M. P., and Tyack, P. L. (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE J. Oceanic Eng.* **28**, 3–12.

King, S. L., Schick, R. S., Donovan, C., Booth, C. G., Burgman, M., Thomas, L., and Harwood, J. (2015). An interim framework for assessing the population consequences of disturbance. *Methods Ecol. Evol.* **6**, 1150–58.

Kleinbaum, D. G., and Klein, M. (2005). Competing risks survival analysis. In *Survival Analysis: A Self-Learning Text*. New York: Springer.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India* **2**, 49–55

Martin, S. W., Martin, C. R., Matsuyama, B. M., and Henderson, E. E. (2015). Minke whales (*Balaenoptera acutorostrata*) respond to navy training. *J. Acoust. Soc. Am.* **137**, 2533–2541.

McHuron, E. A., Schwarz, L. K., Costa, D. P. and Mangel, M. (2018). A state-dependent model for assessing the population consequences of disturbance on income-breeding mammals. *Ecol. Model.* **385**, 133-144.

Miller, P. J. O., Biassoni, N., Samuels, A., and Tyack, P.L. (2000). Whale songs lengthen in response to sonar. *Nature* **405**, 903.

Miller, P. J., 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 Part I: Oceanographic Research Papers*, *56*(7), 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., and Sivle, L. D. (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. *Aquat. Mamm.* **38**, 362–401.

Miller, P. J. O., Antunes, R. N., Wensveen, P. J., Samarra, F. I. P., Alves, A. C., Tyack, P. L., H. Kvadsheim, P., Kleivane, L., Lam, F.P.A., Ainslie, M.A. and Thomas, L. (2014). Dose-response relationships for the onset of avoidance of sonar by free-ranging killer whales. *J. Acoust. Soc. Am.* **135**, 975–993.

Moretti, D., Thomas, L., Marques, T., Harwood, J., Dilley, A, Neales, B., Shaffer, J., McCarthy, E., New, L., Jarvis, S. and Morrissey, R. (2014). A risk function for behavioral disruption of Blainville's beaked whales (*Mesoplodon densirostris*) from mid-frequency active sonar. *PloS ONE*, ***9*****,** p.e85064.

National Academies of Sciences, Engineering and Medicine. (2017). *Approaches to Understanding Cumulative Effects of Stressors of Marine Mammal*. Washington, DC: The National Academies Press.

National Research Council (NRC) (1994). Low-frequency sound in marine mammals: current knowledge and research needs. Washington D.C: The National Academies Press.

National Research Council (NRC) (2005). Marine mammal population and ocean noise: determining when noise causes biologically significant effects. Washington D.C.: The National Academies Press.

Nowacek, D. P., Johnson, M. P., and Tyack, P. L. (2004). North Atlantic right whales (*Eubalaena glacialis*) ignore ships but respond to alerting stimuli. *Proc. Royal Soc. Lond.* ***B*** **271**, 227–231.

Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., LeDuc, C. A., & Hildebrand, J. A. (2007). Behavioral context of call production by eastern North Pacific blue whales. *Mar. Ecol. Prog. Ser.* **330**, 269-284.

Pirotta, E., Mangel, M., Costa, D. P., Mate, B., Goldbogen, J. A., Palacios, D. M.,

Hückstädt, L. A., McHuron, E. A., Schwarz, L. and New, L. (2018). A dynamic state model of migratory behavior and physiology to assess the consequences of environmental variation and anthropogenic disturbance on marine vertebrates. *Am. Nat.* ***191***, E40-E56.

Simon, M., Johnson, M. and Madsen, P. T. (2012). Keeping momentum with a mouthful of water: behavior and kinematics of humpback whale lunge feeding. *J. Exp. Biol.* ***215***, 3786-3798.

Sivle, L. D., Kvadsheim, P. H., Curé, C., Isojunno, S., Wensveen, P. J., Lam, F. P .A., Visser, F., Kleivanec, L., Tyack, P. L., Harris, C. M. and Miller, P. J. O. (2015). Severity of expert-identified behavioural responses of humpback whale, minke whale, and northern bottlenose whale to naval sonar. *Aquat. Mamm*. **41**, 469–502.

Sivle, L. D., Wensveen, P. J., Kvadsheim, P. H., Lam, F. P. A., Visser, F., Curé, C., Harris, C. M., Tyack, P. L. and Miller, P. J. O. (2016). Naval sonar disrupts foraging in humpback whales. *Mar. Ecol. Prog. Ser.* **562,** 211–20.

Southall, B. L. (2017). Noise. In (B. Würsig and H. Thiewesson, Eds.) *Encyclopedia of Marine Mammals, 3rd Edition* (pp. 699-707). Academic Press, New York. ISBN: 9780128043271.

Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene, C. R., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E. and Richardson, W. J. (2007). Marine mammal noise exposure criteria. *Aquat. Mamm*. **33**.

Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene, C.R., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E. and Richardson, W. J. (2009). Marine mammal noise exposure criteria: Initial scientific recommendations. *J. Acoust. Soc. Am.* ***125***, 2517-2517.

Southall, B. L., Moretti, D., Abraham, B., Calambokidis, J., DeRuiter, S. L. and Tyack, P. L. (2012). Marine mammal behavioral response studies in southern California: advances in technology and experimental methods. *Mar. Technol. Soc. J.* ***46***, 48-59.

Southall, B. L., Nowacek, D. P., Miller, P. J. O. and Tyack, P. L. (2016). Experimental field studies to measure behavioral responses of cetaceans to sonar. *Endanger. Species*

Res. ***31***, 293-315.

Therneau, T. M., and Grambsch, P. M. (2000). Multiple events per subject. In *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Therneau, T. M. (2014). A Package for Survival Analysis in S. R Package Version. *R Package Version* 2.37-7.

Tyack, P. L., Zimmer, W. M., Moretti, D., Southall, B. L., Claridge, D. E., Durban, J. W., Clark, C. W., D'Amico, A., DiMarzio, N., Jarvis, S. and McCarthy, E. (2011). Beaked whales respond to simulated and actual navy sonar. *PloS ONE*, ***6***, e17009.

FIGURE CAPTIONS

Figure 1. **Map of overall study area showing locations for all controlled exposure experiments (CEEs) conducted for all (n=42) blue whales**. Treatment types for each CEE (control, simulated mid-frequency active sonar (MFAS), pseudo-random noise (PRN), and real MFAS) are indicated by different symbols.

Figure 2. Movement, diving, and feeding behavior for three CEEs during pre-exposure (baseline), MFAS exposure, and post-exposure phases. Subject movement during each phase is shown in maps (left column) relative to the sound source (black circles) at exposure. Whale diving behavior, lunges (green circles), and received cumulative sound exposure level (cSEL in dB re: $1\mu Pa^2$-s; right axis) are shown in the top panel of plots (right column) showing lunge rate (lunges hr^{-1}), maximum specific acceleration (MSA), heading variance, calculated horizontal speed, and Mahalanobis distance metrics (M. dist. - dashed line indicating maximum value in baseline conditions) are shown in subsequent panels. Corresponding maps and plots are shown for: bw11_210b - CEE #2011-01 (panels A,B); bw11_218b - CEE #2011-06 (panels C,D); and bw13_207a - CEE #2013-06 (panels E,F).

Figure 3. Movement, diving, and feeding behavior for three CEEs for which blue whale prey (krill) schools were measured using active acoustics before and after experimental sequences. Longitudinal plots show individual whale dive profiles (top) and MD plots (bottom) with the exposure phase of CEEs shaded gray. Feeding lunges are marked as green circles and prey patches measured in close horizontal proximity to feeding whales are shown at their respective depth (m) in relative patch density (dB) expressed as relative size and color (denser patches are larger, redder). Corresponding dive profiles and MD plots are shown for: bw11_210b - CEE #2011-01 (panel A); bw11_218b - CEE #2011-06 (panel B); and bw13_207a - CEE #2013-06 (panel C).

Figure 4. Behavioral response probability for deep-feeding blue whales exposed to MFAS and PRN as a function of received cumulative sound exposure level (cSEL in dB re: 1μPa^2-s) for different source-receiver ranges and expert elicitation scored response severities. Response probability model predictions (black lines) with 95% confidence limits (shaded gray areas) are shown for 1, 2, and 5 km source-receiver ranges for moderate (scores 4-6) and high response severity (scores 7-9).

TABLES

Table 1. Controlled exposure experiments (CEEs) conducted for all blue whales in deep-feeding, shallow-feeding, and non-feeding behavioral states. Treatment types for CEEs include: control (no experimental stimuli presented), simulated or real mid-frequency (3-4 kHz) active sonar (MFAS), and pseudo-random noise (PRN) within a similar frequency band (see Southall et al., 2012). Experimental start times are given for 'pre-exposure' (before no-noise control or noise exposure), 'exposure' (during no-noise or noise presentation), and 'post-exposure' (following noise) phases are given in local Pacific Daylight Time (PDT).

Behavioral State at CEE Onset	CEE Type	Subject Identification	CEE Date	CEE Number	Start Times for CEE Phases (local - PDT)		
					Pre-Exposure	*Exposure (min.)*	*Post-Exposure*
Deep-Feeding	CONTROL (n=5)	bw10_241a	8/29/10	2010_07	1125	1155 (30)	1225
		bw10_241_B034	8/29/10	2010_07	1125	1155 (30)	1225
		bw14_212a	7/31/14	2014_02	1346	1416 (30)	1446
		bw14_213a	8/1/14	2014_03	1506	1536 (30)	1606
		bw14_251a	9/8/14	2014_05	1155	1225 (30)	1255
Deep-Feeding	MFAS (n=13)	bw10_239b	8/27/10	2010_05	1204	1234 (30)	1304
		bw10_246a	9/3/10	2010_12	1323	1353 (25)[a]	1418
		bw10_246b	9/3/10	2010_12	1323	1353 (25)[a]	1418
		bw11_210a	7/29/11	2011_01	1455	1525 (30)	1555
		bw11_210b	7/29/11	2011_01	1455	1525 (30)	1555
		bw11_213b	8/1/11	2011_03	1216	1246 (30)	1316
		bw11_219b	8/7/11	2011_07	1728	1758 (24)[b]	1822
		bw11_220b	8/8/11	2011_08	1519	1549 (30)	1619
		bw13_191a	7/10/13	2013_03	1219	1319 (58)[c]	1417
		bw14_211b	7/30/14	2014_01	1524	1554 (30)	1624
		bw14_218a	8/6/14	2014_04	1131	1201 (30)	1231
		bw14_256a	9/13/14	2014_07	1015	1045 (30)	1115
		bw14_262b	9/19/14	2014_10	1032	1102 (28)[a]	1130
Deep-Feeding	PRN (n=11)	bw10_243a	8/31/10	2010_09	1209	1239 (30)	1309
		bw10_243b	8/31/10	2010_09	1209	1239 (30)	1309
		bw10_244b	9/1/10	2010_10	1654	1724 (30)	1754
		bw10_244c	9/1/10	2010_10	1654	1724 (30)	1754
		bw10_245a	9/2/10	2010_11	1322	1352 (30)	1422
		bw10_266a	9/23/10	2010_19	1559	1629 (30)	1659
		bw11_211a	7/30/11	2011_02	1038	1108 (18)[a]	1126
		bw11_214b	8/2/11	2011_04	1050	1120 (30)	1150
		bw11_218b	8/6/11	2011_06	1709	1739 (23)[b]	1802
		bw11_221a	8/9/11	2011_09	1429	1459 (30)	1529
		bw11_221b	8/9/11	2011_09	1429	1459 (30)	1529
Shallow-Feeding	CONTROL (n=1)	bw13_207a	7/26/13	2013_06	1714	1744 (30)	1814
Shallow-Feeding	MFAS (n=7)	bw10_235a	8/23/10	2010_01	1117	1147 (30)	1217
		bw10_235b	8/23/10	2010_01	1117	1147 (30)	1217
		bw10_238a	8/26/10	2010_04	1143	1213 (30)	1243
		bw10_240a	8/28/10	2010_06	0917	0947 (30)	1017
		bw10_240b	8/28/10	2010_06	0917	0947 (30)	1017
		bw13_259a	9/16/13	2013_16	1046	1116 (30)	1146
		bw14_262a	9/19/14	2014_10	1032	1102 (28)[a]	1130
Non-Feeding	MFAS (n=2)	bw10_235_B019	8/23/10	2010_02	1617	1647 (18)[a]	1705
		bw10_265a	9/22/10	2010_17	1252	1322 (19)[b]	1341
Non-Feeding	PRN (n=3)	bw10_251a	9/8/10	2010_16	1450	1520 (30)	1550
		bw11_218a	8/6/11	2011_06	1709	1739 (23)[b]	1802
		bw12_292a	10/18/12	2012_05	1304	1334 (30)	1404

[a] Required source shut-down prior to full duration because individuals of non-focal species (California sea lions (*Zalophus californianus*)) entered mandated source shut-down zone.

[b] Required source shut-down prior to full duration because individuals of non-focal species (either bottlenose dolphins (*Tursiops truncatus*) or common dolphins (*Delphinus delphis*)) entered mandated source shut-down zone.

[c] Longer specified pre-exposure, exposure, and post-exposure period for operational Navy 53C sonar.

Table 2. Controlled exposure experiment (CEE) results for all blue whales in deep-feeding, shallow-feeding, and non-feeding behavioral states. Maximum received cumulative sound exposure levels (cSEL; dB re: 1µPa^2-s) are given for all individuals for all CEEs involving noise exposure. Behavioral changes identified using with Mahalanobis distance statistical change-point methods and expert evaluation scoring (see text) are presented for each whale and summarized within each behavioral state and CEE treatment type. Relative confidence (low, med, high) for expert scoring panels as well as the highest attributed response severity are provided. Where behavioral changes were detected, received cSEL is given at change-points identified by MD and expert scoring methods (see text). Whether analytical methods agree in detecting changes is identified and total changes for MD analyses (excluding instances where changes were associated with feeding onset) and expert scoring results are compared within categories.

Subject Behavioral State and CEE Type	Subject ID	Source-Whale Range (km)	Max. cSEL (dB re: 1μPa²-s)	Behavioral Change Identified? — *Mahalanobis Distance (MD)*		Behavioral Change Identified? — *Expert Scoring (ES)*			Methods Agree?	Total MD Changes**	Total ES Changes
				Change Identified ?	Received cSEL at change point	Change? (confidence)	Scored severity	Received cSEL at change point			
Deep-Feeding CONTROL (5)	bw10_241a	0.3	n/a	NO	n/a	NO (high)	0	n/a	YES	1 of 5	0 of 5
	bw10_241_B034	1.75	n/a	NO	n/a	NO (low)	0	n/a	YES		
	bw14_212a	1.3	n/a	YES	n/a	NO (low)	0	n/a	NO		
	bw14_213a	0.7	n/a	NO	n/a	NO (low)	0	n/a	YES		
	bw14_251a	1.25	n/a	YES*	n/a	NO (high)	0	n/a	YES		
Deep-Feeding MFAS (13)	bw10_239b	2.8	164	YES	137	YES (mod)	5	128	YES	9 of 13	8 of 13
	bw10_246a	1.45	169	NO	-	NO (mod)	0	-	YES		
	bw10_246b	1.3	169	YES	150	NO (high)	0	-	NO		
	bw11_210a	1.2	167	YES	165	NO (mod)	0	-	NO		
	bw11_210b	0.8	171	YES	119	YES (high)	7	119	YES		
	bw11_213b	1.0	169	YES	113	NO (high)	0	-	NO		
	bw11_219b	1.25	162	YES*	-	YES (mod)	4	155	NO		
	bw11_220b	1.15	142	YES	140	YES (high)	5	125	YES		
	bw13_191a	19.5	153	YES*	-	NO (high)	0	-	YES		
	bw14_211b	0.7	149	YES	140	YES (mod)	5	138	YES		
	bw14_218a	1.1	132	NO	-	YES (low)	5	116	NO		
	bw14_256a	0.8	154	YES	120	YES (high)	7	114	YES		
	bw14_262b	1.4	145	YES	141	YES (low)	3	125	YES		
Deep-Feeding PRN (11)	bw10_243a	4.6	157	NO	-	NO (high)	0	-	YES	5 of 11	4 of 11
	bw10_243b	0.8	160	YES*	-	NO (low)	0	-	YES		
	bw10_244b	1.15	168	YES	105	NO (mod)	0	-	NO		
	bw10_244c	1.6	160	YES	158	YES (high)	7	110	YES		
	bw10_245a	7.7	149	YES*	-	NO (high)	0	-	YES		
	bw10_266a	1.25	160	YES	148	YES (high)	7	148	YES		
	bw11_211a	1.1	162	NO	-	NO (high)	0	-	YES		
	bw11_214b	0.4	160	YES	109	YES (high)	6	109	YES		
	bw11_218b	1.2	168	NO	-	NO (high)	0	-	YES		
	bw11_221a	0.6	160	YES	124	YES (low)	5	97	YES		
	bw11_221b	0.6	162	NO	-	NO (low)	0	-	YES		
Shallow-Feeding CONTROL (1)	bw13_207a	0.5	n/a	YES*	n/a	NO (mod)	0	n/a	YES	0 of 1	0 of 1
Shallow-Feeding MFAS (7)	bw10_235a	1.05	170	NO	-	NO (low)	0	-	YES	0 of 7	0 of 7
	bw10_235b	1.7	145	YES*	-	NO (mod)	0	-	YES		
	bw10_238a	0.45	152	NO	-	NO (mod)	0	-	YES		
	bw10_240a	0.5	169	YES*	-	NO (high)	0	-	YES		
	bw10_240b	3.7	165	NO	-	NO (high)	0	-	YES		
	bw13_259a	5.2	134	NO	-	NO (mod)	0	-	YES		
	bw14_262a	1.4	142	YES*	-	NO (mod)	0	-	YES		
Non-Feeding MFAS (2)	bw10_235_B019	1.9	158	YES	108	YES (high)	7	108	YES	2 of 2	1 of 2
	bw10_265a	1.9	158	YES	148	NO (mod)	0	-	NO		
Non-Feeding PRN (3)	bw10_251a	0.85	159	YES	123	YES (low)	7	102	YES	3 of 3	1 of 3
	bw11_218a	5.6	137	YES	102	NO (mod)	0	-	NO		
	bw12_292a	1.15	157	YES	127	NO (high)	0	-	NO		

* Associated with onset of feeding in MD change-point; this was not scored as a response within expert scoring

** Not including identified MD changes associated with feeding onset