

Severity of Expert-Identified Behavioural Responses of Humpback Whale, Minke Whale, and Northern Bottlenose Whale to Naval Sonar

Lise D. Sivle,¹ Petter H. Kvadsheim,² Charlotte Curé,⁷ Saana Isojunno,³ Paul J. Wensveen,³ Frans-Peter A. Lam,⁴ Fleur Visser,^{5,6} Lars Kleivane,² Peter L. Tyack,³ Catriona M. Harris,⁸ and Patrick J. O. Miller³

¹Institute of Marine Research (IMR), PO Box 1870, Nordnes, NO-5817, Bergen, Norway

E-mail: lise.doksaeter.sivle@imr.no

²Norwegian Defence Research Establishment (FFI), NO-3191, Horten, Norway

³Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, Fife KY16 8LB, UK

⁴Netherlands Organisation for Applied Scientific Research (TNO), PO Box 96864, The Hague, 2509 JG, The Netherlands

⁵Kelp Marine Research (KMR), 1624 CJ, Hoorn, The Netherlands

⁶Behavioural Biology Group, Leiden University, PO Box 9505, 2300 RA, Leiden, The Netherlands

⁷Cerema, Dter Est, Acoustics Group, F-67035, Strasbourg Cedex 2, France

⁸Centre for Research into Ecological and Environmental Modeling, Buchanan Gardens, University of St Andrews, St Andrews, Fife, KY16 9LZ, UK

Abstract

Controlled exposure experiments using 1 to 2 kHz sonar signals were conducted with 11 humpback whales (*Megaptera novaeangliae*), one minke whale (*Balaenoptera acutorostrata*), and one northern bottlenose whale (*Hyperoodon ampullatus*) during three field trials from 2011 to 2013. Ship approaches without sonar transmissions, playbacks of killer whale vocalizations, and broadband noise were conducted as controls. Behavioural parameters such as horizontal movement, diving, social interactions, and vocalizations were recorded by animal-attached tags and by visual and acoustic tracking. Based on these data, two expert panels independently scored the severity of behavioural changes that were judged likely to be responses to the experimental stimuli, using a severity scale ranging from no effect (0) to high potential to affect vital rates (9) if exposed repeatedly. After scoring, consensus was reached with a third-party moderator. In humpback whales, killer whale playbacks induced more severe responses than sonar exposure, and both sonar exposures and killer whale playbacks induced more responses and responses of higher severity than the no-sonar ship approaches and broadband noise playbacks. The most common response during sonar exposures in all three species was avoidance of the sound source. The most severe responses to sonar (severity 8) were progressive high-speed avoidance by the minke whale and long-term area avoidance by the bottlenose whale. Other severe

responses included prolonged avoidance and cessation of feeding (severity 7). The minke whale and bottlenose whale started avoiding the source at a received sound pressure level (SPL) of 146 and 130 dB re 1 μ Pa, respectively. Humpback whales generally had less severe responses that were triggered at higher received levels. The probability of severity scores with the potential to affect vital rates increased with increasing sound exposure level (SEL_{cum}). The single experiments with minke and bottlenose whales suggest they have greater susceptibility to sonar disturbance than humpback whales, but additional studies are needed to confirm this result.

Key Words: behavioural response, naval sonar, severity scoring, controlled exposure experiment, arctic biology, humpback whale, *Megaptera novaeangliae*, minke whale, *Balaenoptera acutorostrata*, northern bottlenose whale, *Hyperoodon ampullatus*

Introduction

All marine mammals use sound to gather information about their surroundings, either through active echolocation (Au, 1993) or passive listening (Gannon et al., 2005). Sound has a fundamental role in prey and predator detection, animal communication, and navigation (Richardson et al., 1995; Tyack, 2000; Simpson et al., 2005). During the last century, there has been a significant overall increase of underwater anthropogenic noise, leading to

increased background noise levels and more frequent exposure to high intensity sounds (McDonald et al., 2008). Exposure to noise might lead to direct physical injuries (Southall et al., 2007) or behavioural changes (e.g., Morton & Symonds, 2002; Tyack et al., 2011; DeRuiter et al., 2013; Kuningas et al., 2013; Miller et al., 2015). Among the contributors to increased noise levels are navies with modern active long-range sonars, typically operating in the frequency range of 0.2 to 10 kHz, which is within the range of hearing sensitivity for most marine mammals (Popper & Ketten, 2008).

Potential negative effects of such sonars on marine mammals have received particular attention after several atypical cetacean mass strandings in connection with naval sonar activity, mostly (but not exclusively) involving beaked whales (Frantzis, 1998; Balcomb & Claridge, 2001; D'Amico et al., 2009). These stranding events might be directly or indirectly caused by behavioural responses (Cox et al., 2006). Besides stranding, other negative effects on vital rates might also be induced by behavioural responses. Behavioural effects of sonar can range from subtle effects such as short changes in vocal behaviour (Miller et al., 2000; Croll et al., 2001; Fristrup et al., 2003; Alves et al., 2014) and dive patterns (Sivle et al., 2012) to more severe responses such as habitat avoidance (Tyack et al., 2011; DeRuiter et al., 2013; Miller et al., 2015), typically also associated with cessation of feeding (Miller et al., 2012; Goldbogen et al., 2013; Isojunno & Miller, 2015) and even separation from dependent offspring (Miller et al., 2012). The specific context of the exposure event appears to alter the likelihood of responses (Goldbogen et al., 2013) and may also explain high variability in response thresholds (Antunes et al., 2014; Miller et al., 2015). Eventually, the behavioural context of exposed animals will probably also determine the biological significance of responses (Southall et al., 2007; Ellison et al., 2012; Miller et al., 2012).

Experimental exposure of free-ranging individual animals to high intensity sounds such as naval sonar or seismic signals is one approach for providing a better understanding of the sound-level thresholds at which responses occur, the context of the reactions, and their potential biological relevance on populations in the wild (Tyack et al., 2004; Tyack, 2009). Such experiments are designed to control the species and location, timing of the exposure, and target received sound levels, as well as to allow for the collection of appropriate data on natural baseline behaviour. Baseline behaviour of an individual is compared to its behaviour during sound exposure and, thus, is used to evaluate the likelihood of a behavioural response at different received levels.

This experimental approach also allows testing of various control stimuli. Indeed, comparison of responses to an approaching ship transmitting sonar vs a ship with no transmissions is used to separate the effects of sonar from the effect of the source ship itself (e.g., Miller et al., 2012). Furthermore, comparison of responses to sonar with natural anti-predator responses can help in understanding the biological significance of observed behavioural responses to anthropogenic disturbances (Frid & Dill, 2002).

Southall et al. (2007) provided a useful response severity scale for evaluating observed changes in animal behaviour in response to anthropogenic noise. Using that scale, Miller et al. (2012) scored behavioural responses collected from high-resolution tags and visual observations of long-finned pilot whales (*Globicephala melas*), killer whales (*Orcinus orca*), and sperm whales (*Physeter macrocephalus*) exposed to naval sonar and control signals. They identified some behavioural responses that were not considered in the original scale (Southall et al., 2007), and they made a slightly modified scale that still followed the original logic (Miller et al., 2012). Moreover, Miller et al. (2012) found marked differences in the severity of responses across species and signal frequency. In the present study, we have used this modified severity scale to determine the severity of expert-scored behavioural responses of one odontocete species, the northern bottlenose whale (*Hyperoodon ampullatus*), and two mysticete species, the humpback whale (*Megaptera novaeangliae*) and the minke whale (*Balaenoptera acutorostrata*), to controlled exposure of naval sonar signals, no-sonar ship approaches, and playbacks of killer whale sounds and broadband noise.

Methods

Animal Welfare Considerations

All animal research activities were permitted by the Norwegian Animal Research Authority (Permit No. S-2011/38782) and were approved by the Animal Welfare Ethics Committee at the University of St Andrews and the Institutional Animal Care and Use Committee of the Woods Hole Oceanographic Institution. All of our experiments followed a safety plan designed to protect the welfare of the study animals as well as other animals in the area. Visual observers continuously scanned for whales throughout the exposures with a detailed plan in place to stop sonar transmissions if potentially hazardous responses occurred or if any animal came too close to the sonar source. Our experimental design further involved limited duration exposure periods, changing of subjects between experiments, and exposure of a limited

number of animals, reducing risk of harm to experimental subjects.

Experimental Design and Protocol

Sonar exposure experiments were conducted in the Arctic Atlantic Ocean near Bear Island and Svalbard and off Jan Mayen during 2011, 2012, and 2013 using the Norwegian Defence Research Establishment (FFI) research vessel *HU Sverdrup II* (“*HUS*”). Details of each annual cruise can be found in specific cruise reports (Kvadsheim et al., 2011, 2012, 2014). Details of the experimental procedures can be found in Kvadsheim et al. (2015) and are only summarized here. The protocol consisted of five phases: (1) searching, (2) tagging, (3) baseline pre-exposure, (4) experimental exposure, and (5) post-exposure (Kvadsheim et al., 2015). Sonar signals were transmitted from *HUS*, whereas killer whale sounds and broadband noise playbacks were broadcasted from a small boat. The whale’s position and social and surface behaviour were recorded when the whale surfaced repeatedly at more than 2 min apart, using the procedure of Visser et al. (2014).

The position of each tagged whale was estimated in real-time by the observers and relayed to the sonar operator on the source vessel (*HUS*). Sonar transmissions were always initiated by a ramp-up procedure, which implied a gradual increase of source level from 152 to 214 dB re 1 μ Pa m. The primary goal of movements of the source vessel with respect to the subject whale was to achieve a gradual escalation of sonar received levels, but the ship’s movements during sonar transmissions were different between the three species:

- *Humpback Whales*—Three different experimental conditions were used for humpback whales: (1) no-sonar-control approaches, (2) ramp-up sonar exposures, and (3) no-ramp-up sonar exposures, each with a 10-min duration. During the ramp-up exposures, sonar transmission was initiated at approximately 1.3-km distance from the tagged animal, and the source ship approached with a speed of 4.1 m/s on a predetermined straight course to intercept the animal’s course (determined prior to the start of the exposure), while transmitting a 5-min ramp-up. Full power transmission then continued for another 5 min. The no-sonar control approach and the no-ramp-up approach followed the exact same procedure, except that no sonar pulses were transmitted during no-sonar controls, and transmissions only started at full power 5 min into the session during no-ramp-up exposures. This protocol was specifically designed to partly test the efficacy of ramp up

as a mitigation procedure, as well as to assess behavioural responses to the sonar exposure.

- *Minke Whale* – *HUS* approached the whale from a distance of 8 km at 4.4 m/s, first doing a no-sonar run, followed by a sonar exposure run with a 10-min ramp-up. The animal was approached to intercept its course, which was determined prior to start of the exposure. The sonar exposure continued until 5 min after the closest point of approach (CPA), but sonar exposure never lasted longer than 60 min. This protocol was identical to that of Miller et al. (2012), which allows comparison of results.
- *Northern Bottlenose Whale* – This species can dive for more than 1 h (Hooker & Baird, 1999). Good tracking of these animals, therefore, relies on acoustic tracking during deep dives in combination with visual tracking when they are at the surface. To achieve this goal, the source vessel sailed 2 km $\times$ 2 km boxes, towing an acoustic array at 100 to 200 m depth around the estimated position of the whale throughout the tracking period. During sonar exposure, the source ship started $\sim$ 5 km from the position of the whale and sailed a predetermined track consisting of $\frac{3}{4}$ of a 2 km $\times$ 2 km box at 2.6 m/s while first transmitting a 20-min ramp-up, followed by 15 min of full power transmissions (Miller et al., 2015). During exposure, the focal whale was not inside the box but several km away from it, and, thus, the angle from the whale to the source vessel changed very little. This protocol was similar to those of Tyack et al. (2011) and DeRuiter et al. (2013) to ensure comparability with existing data on other beaked whales.

Playbacks of Killer Whale Sound and Broadband Noise

These were only conducted with humpback whales (Curé et al., 2015). The killer whale sound playback was used as a positive control for assessing responses of animals to a natural threatening stimulus (i.e., predation risk), and the noise playback was used as a negative control of the killer whale sound playback to test the animal’s reaction to any unspecific noise.

The killer whale playback stimulus was previously recorded in the North Pacific using Dtags (Miller et al., 2010) and corresponded to natural sequences of vocalizations of a killer whale group attacking and feeding on prey. The stimulus thus represents “unfamiliar” mammal-eating killer whale sounds expected to be perceived as an immediate risk of predation (Deecke et al., 2002; Curé et al., 2013, 2015). The control stimulus was

a broadband noise playback (0.5 to 10 kHz). Both the killer whale sound and the broadband noise were of 15-min duration and had similar root mean square (rms) source levels (146 to 152 dB re 1 $\mu\text{Pa m}$), corresponding to typical source levels of killer whale sounds (Miller, 2006; Curé et al., 2013) but much lower than the source level of the sonar (214 dB re 1 $\mu\text{Pa m}$). Playbacks were conducted from a stationary small boat, and further details of the playback protocol are given in Curé et al. (2012, 2013, 2015).

Sonar and No-Sonar Control Exposures

The sonar source (Socrates II, TNO, The Netherlands) deployed from the *HUS* during the experiments was a multipurpose towed acoustic source developed for underwater acoustic research and previously tested as a prototype sonar system on operational Royal Netherlands Navy frigates. The transmitted signals were 1.0 to 2.0 kHz frequency-modulated hyperbolic upsweeps in the bottlenose whale experiment, and 1.3 to 2.0 kHz frequency-modulated hyperbolic upsweeps in the humpback and minke whale experiments. The ramp-up procedure was used as part of the risk mitigation plan, and also as part of the experimental design that aimed to gradually escalate the acoustic dose. The sonar pulse repetition interval for all experiments was 20 s, and the signal duration was 1 s, except during the ramp-up period of humpback whale experiments for which the signal duration was 0.5 s. Humpback whales and the minke whale were also subject to the same approach by the source vessel but without sonar transmissions (no-sonar control) to enable comparison of potential responses to the sonar with responses to the approaching source vessel alone.

Data-Recording Tags

Whales were tagged with movement-and-sound-recording Dtags (Johnson & Tyack, 2003) or Ctags (Kvadsheim et al., 2011). The Dtag was attached to the whale with suction cups using a 15-m cantilevered carbon fibre pole or a pneumatic remote deployment system (ARTS; Kvadsheim et al., 2009) and stayed on for 16 to 18 h. The Dtags had one or two hydrophones and recorded sound with 16-bit resolution at a sampling rate of 96 kHz, and were also equipped with depth and a three-dimensional (3D) accelerometer, and magnetometer sensors sampled at 50 Hz. Additionally, the Dtag contained a VHF transmitter and a GPS data logger (SirTrack, F2G 134A, Fastloc 2), which recorded the horizontal location of the tagged whale.

The CTAG was used for the minke whale experiment because previous experiences with minke whales confirmed that suction cup attachment

was unreliable (Kvadsheim et al., 2011). The Ctag was deployed using the ARTS system and was attached to the whale by a 5-cm-long barb that penetrates the skin and anchors within the blubber. It contained a VHF-transmitter and a Star Odditime-depth recorder (Kvadsheim et al., 2011). The Ctag was released from the animal using a galvanic time release after 19 h.

Visual and Acoustic Observations and Tracking

Visual searching and tracking was conducted from an elevated marine mammal observation platform on the *HUS*. Focal follows of the group containing the tagged whale were either made from the *HUS* (bottlenose whales) or from an independent small boat deployed from the *HUS* (humpback and minke whales). Both platforms had a digital VHF direction finder (DFHorten; LK-ARTS, Norway) used in combination with visual observers. Passive acoustic monitoring was used to locate bottlenose whales and track them during experiments (Kvadsheim et al., 2014), using a dedicated hydrophone array developed by TNO (von Benda-Beckmann et al., 2010). Surface behaviour of the focal group was recorded systematically from the first sighting to the end of the experiment following Visser et al. (2014).

Processing and Analysis of Movement Data

For humpback and minke whales, horizontal tracks of the tagged animals were created from the GPS data and visual track data. A dead-reckoned track ("pseudotrack") based on the output of the tag's orientation and acoustic sensors was produced for the tagged bottlenose whale (Miller et al., 2015) as this animal was lost for visual tracking, and the tag did not record GPS data.

The horizontal speed and direction of movement of the tagged whale were calculated from the horizontal locations obtained from sightings at the surface and/or by GPS logger, or from the dead-reckoned track. For sightings and GPS, speed was calculated as the total great circle distance travelled over three surfacings divided by the total time between them. Directness was calculated by dividing distance covered by the whale during three observation points of a track segment (i.e., the distance between the first surfacing and the third surfacing) by the cumulative distance between all three points and is given as values between 0 (circular movement) and 1 (straight movement).

Acoustic recordings of tagged whale sounds were analyzed only for the bottlenose whale. There was no acoustic sensor on the Ctag, and, thus, no acoustic data were collected for the minke whale. For humpback whales, an extensive scan of the acoustic recordings on the tags revealed that they

made very few sounds, and we did not pursue this any further to reduce time costs. The audio recording of the bottlenose whale was audited to identify foraging sounds produced by the tagged whale and other nearby whales (Miller et al., 2015). Click and buzz sounds were used as acoustic cues indicating foraging activity. Audio files recorded by the Dtags were displayed as spectrograms (Blackman-Harris window; FFT length: 512) with a 15-s duration window. The start and end of each detected sound was identified and marked, and it was ascribed to the tagged whale or another whale depending upon its relative amplitude and spectral characteristics. Other biological sounds from the tagged or nearby whales were also annotated. Details of the method can be found in Miller et al. (2015).

For humpback whales, the acoustic record of the Dtag was used to detect lunge feeding events. Humpback lunge feeding involves engulfing a large volume of prey-rich water in the flexible buccal cavity and filtering out prey with the baleen. A lunge is characterized by an increase in speed followed by an abrupt drop in speed as the whale first accelerates forward and then slows down quickly after the jaw opens. The low-frequency flow noise measured on the acoustic record of the Dtag is a useful proxy for the whale's speed through the water and was used here to identify lunge events, replicating the method of Simon et al. (2012). We developed an automatic lunge detector that identified events with noise peaks that exceed the 90th percentile (for depth > 5 m) of the flow noise (< 500 Hz) in all dives deeper than 5 m and which were followed by at least a 12 dB drop in flow noise within 5 s. This 5-s period was truncated if the whale reached the surface (depth < 0.5 m) to avoid drops in the noise when the whale surfaced to breathe.

Processing and Analysis of Received Sonar Signals

Sonar signals recorded by the Dtag were extracted for detailed analysis of the sonar received levels (Miller et al., 2011; Kvadsheim et al., 2015). For each sonar pulse, we measured the broadband maximum rms sound pressure level (SPL) over a 200-ms averaging window and the broadband cumulative sound exposure level (SEL_{cum}) throughout the exposure session. These broadband received levels were calculated from $1/3$ -octave bands between 1 to 40 kHz in which the signal exceeded the noise on the tag by 10 dB or more. The levels of some sonar pulses during the ramp-up at the start of each exposure session were below this threshold. The received level of those pulses was calculated by extrapolation using the measured level of the closest ping in time adjusted

for the known difference in source level. The acoustic sensitivity of the Dtags, determined from calibration measurements, was (mean $\pm$ SD) -185 ± 4 dB re $1 \mu\text{Pa}^{-1}$ ($N = 6$ tags).

The Ctag deployed on the minke whale did not have any acoustic sensors, so the received SPLs were estimated from acoustic propagation modelling. The ray-trace model LYBIN (Dombestein & Gjersøe, 2012) was used to calculate acoustic propagation loss in two dimensions (range and depth) using the transmission characteristics of the Socrates source. The model was re-run with two different sound speed profiles measured in the exposure area shortly after the sonar exposure (Kvadsheim et al., 2015) and with two different source depths (source depths varied between 60 and 69 m during the exposure), and the mean value was used (average standard deviation was 3.6 dB). The depth of the tagged whale was calculated from the tag's pressure sensor, and the horizontal movement under water was interpolated between every surfacing, resulting in a horizontal and vertical position of the whale at the time of each sonar ping. Both the GPS position and depth of the sonar source were recorded, and, thus, the horizontal distance between the whale and the source for each ping could be estimated. Received SPLs were calculated as transmitted source level minus mean propagation loss estimated by LYBIN. The received level estimates were found to be within 5 dB of levels measured by calibrated hydrophones towed by the small boat tracking the whale (Kvadsheim et al., 2015).

Scoring Severity of Expert-Identified Behavioural Responses

Based upon an extensive literature review on behavioural effects of sound on marine mammals, a panel of experts in marine mammal behaviour scaled documented behavioural responses according to the putative severity for the affected animals (Southall et al., 2007). The scale ranged from no response (0), responses that are not likely to affect vital rates (1 to 3), responses that have the potential to affect vital rates (4 to 6), to responses that are likely to affect vital rates (7 to 9) if the exposure is of sufficient duration or repeated. The severity of a response depends upon both the type of behavioural change and the duration of the response (Southall et al., 2007).

Each experiment of the present study was represented by a series of standardized data plots; a geographic track plot (GIS plot) and a time series data plot were provided for the whole tag record, with close-ups of each pre-exposure baseline and exposure periods (see Figure 1 through 3 for examples of plots for exposure period). Full sets of plots are available in Kvadsheim et al. (2015).

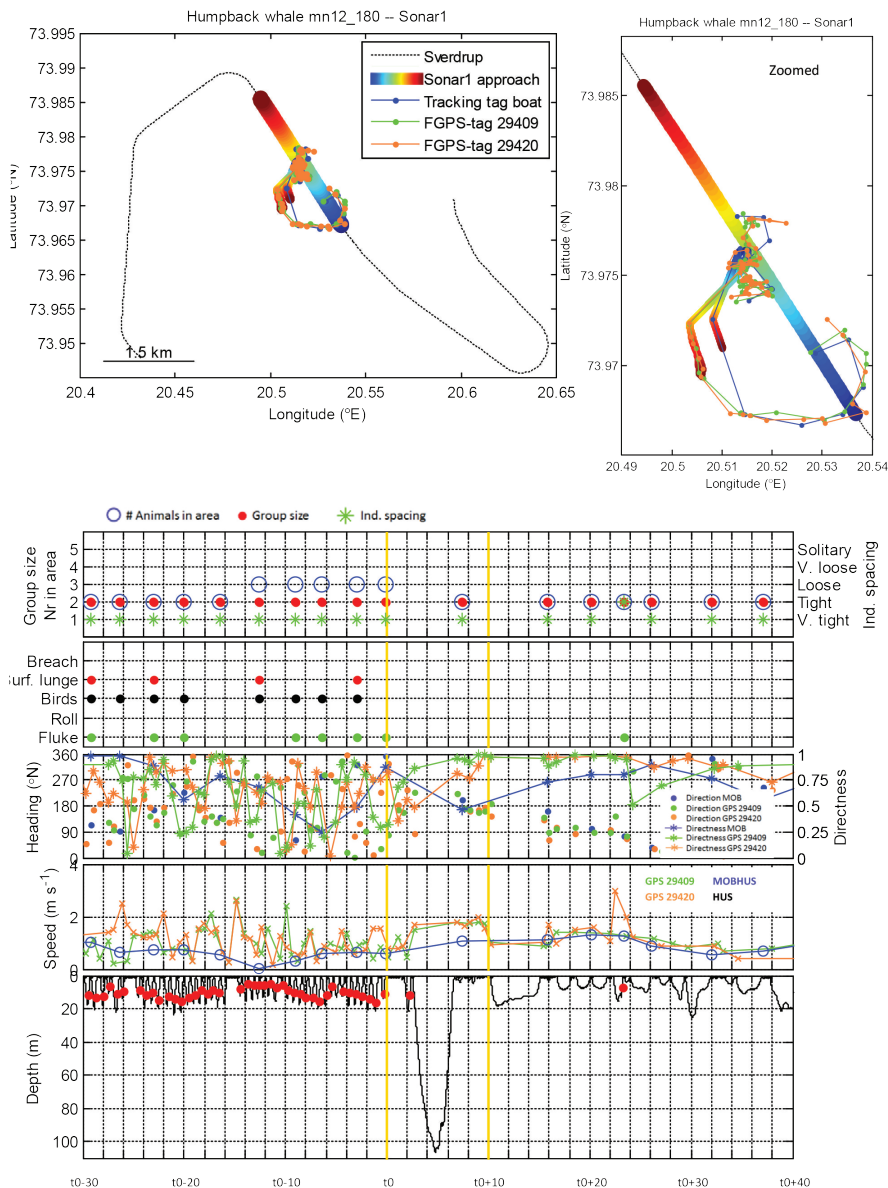


Figure 1. Plot for humpback whale (*Megaptera novaeangliae*) (mn12_180ab) during the time period from 30 min prior to the first exposure session (sonar 1) to 30 min after the session. Upper panels show the track of the source boat (*HUS*) and the whale, with the upper right panel showing a zoomed-in view around the exposure area. The colour coding on the animal and ship tracks indicates the time since the start of the exposure session, with blue and red marking the start and end, respectively. Tracks of the whale are based upon the visual tracking, and the GPS loggers on the whale are plotted, with each dot on the track representing a position fix. Mn12_180ab was tagged with two Dtags, and both GPS tracks are shown. The lower panel shows time series plots of different variables recorded. From the top: group structure variables (e.g., group size, number of individuals in the area, and individual spacing), surface display variables (e.g., observed breaches, surface lunges, rolls, flukes, and if there were birds surrounding the focal animal[s]), heading and directness (i.e., a measure of how straight the animal is swimming: 1 = straight line, 0 = circular) measured from the visual observations and GPS, speed of the animal based upon the visual track and GPS tracks, and dive profile (black) overlaid with detected feeding lunges marked as red dots. Vertical yellow lines indicate the start and end of the sonar exposure period. More plots of this animal, including the entire time series, can be found in Kvadsheim et al. (2015).

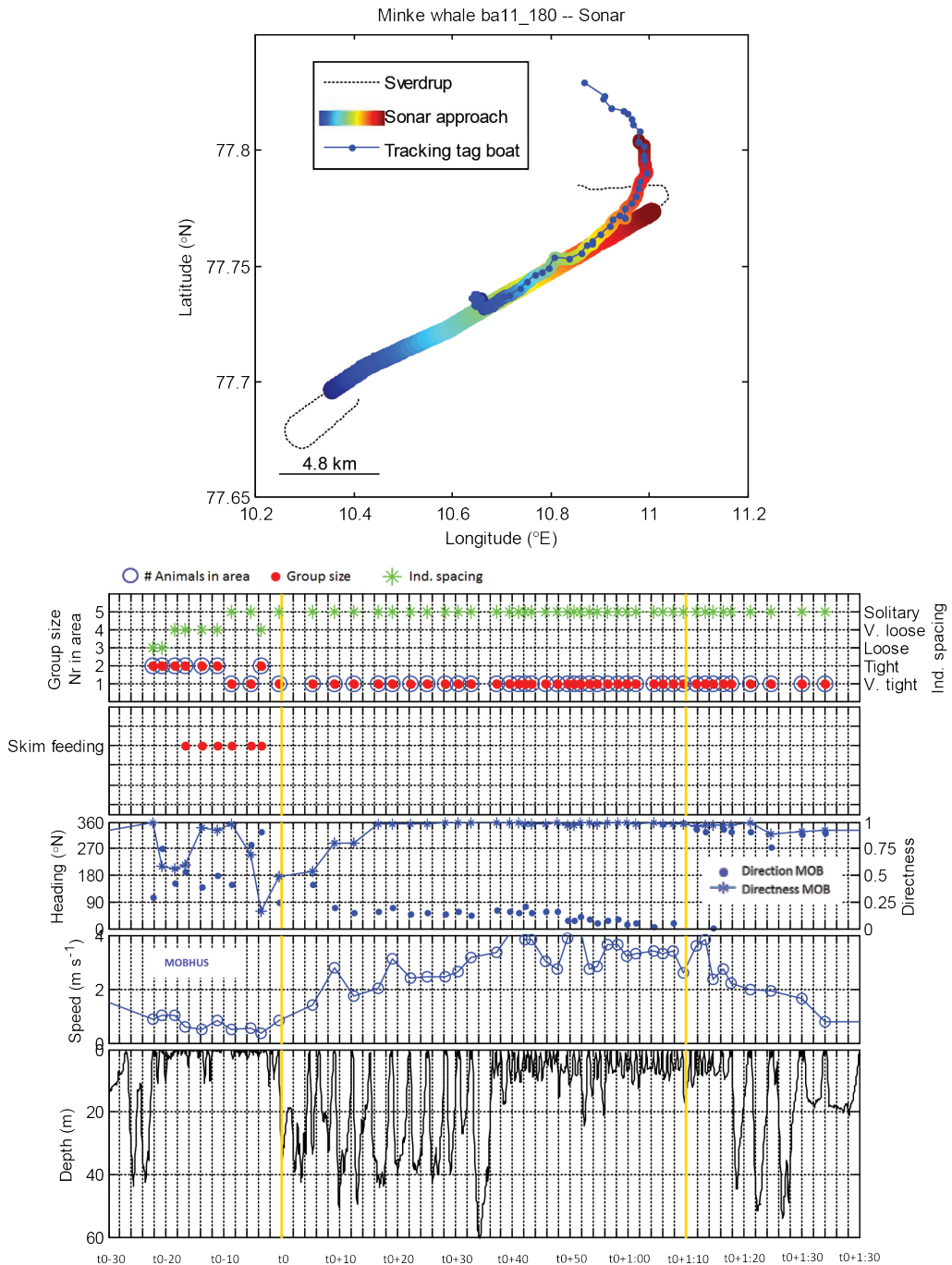


Figure 2. Plot for minke whale (*Balaenoptera acutorostrata*) (ba11_180) during the time period from 30 min prior to exposure to 30 min after sonar exposure to sonar 1. Format of figure follows Figure 1. The minke whale did not carry a GPS logger and, therefore, only the visual track is shown. Each dot on the track represents a visual position fix. The sensors on the Ctag used on the minke whale did not allow detection of lunges, so lunges are not indicated. However, skim-feeding at the surface was observed, and the timing of these are indicated in the figure. More plots of this animal, including the entire time series, can be found in Kvadsheim et al. (2015).

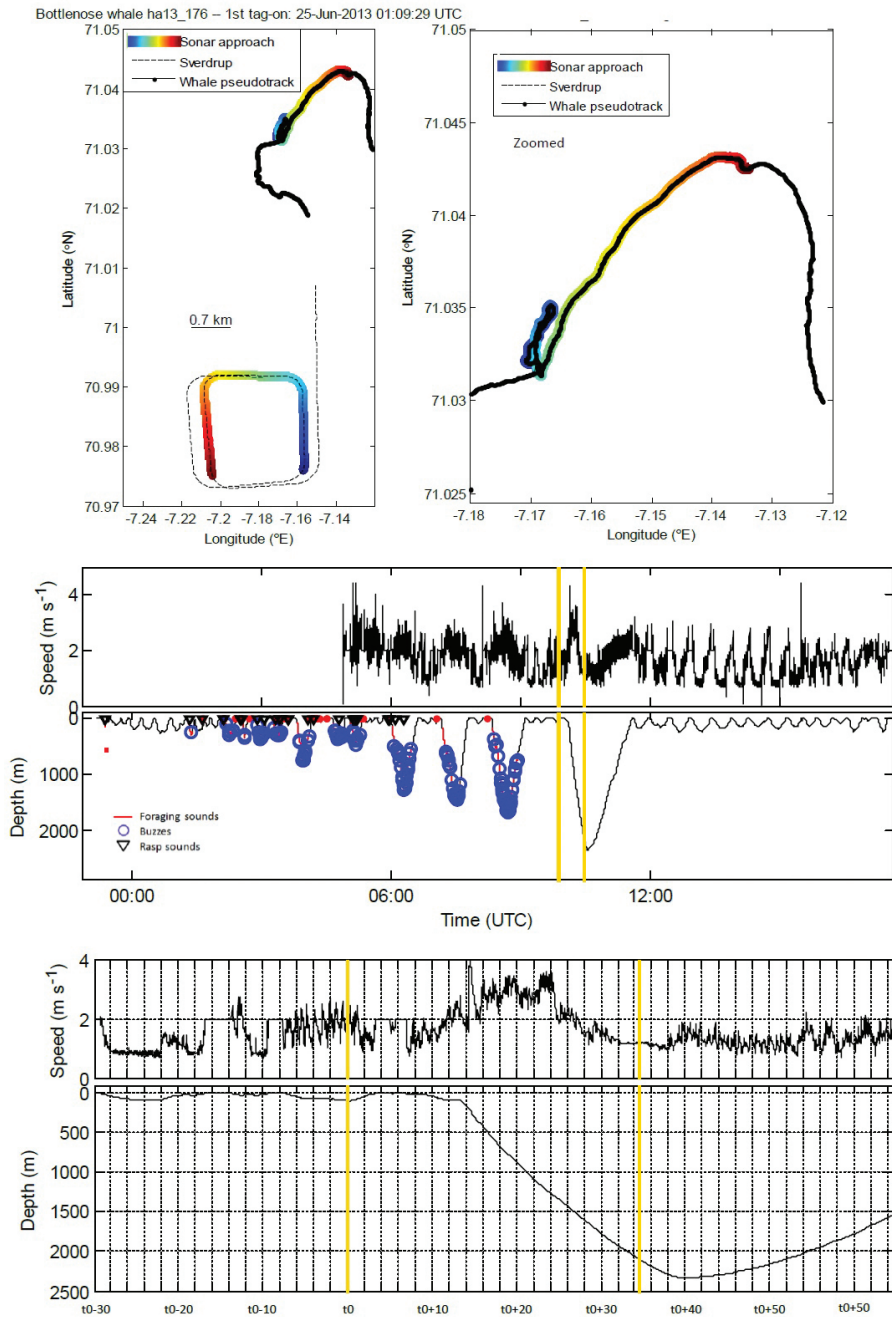


Figure 3. Plot for northern bottlenose whale (*Hyperoodon ampullatus*) (ha13_176). Upper panel is horizontal track. The animal was lost for visual tracking and, therefore, the track of the whale is a reconstructed dead-reckoned track (pseudotrack) based upon estimated speed, pitch, and heading recorded by the tag. The middle and lower panel show time series plots of the estimated speed based on the flow noise on the tag and the dive profile (black) with identified vocalizations indicated. The middle panel shows the entire time series, while the lower panel is a zoomed-in view of the time plots showing the time period from 30 min prior to exposure to 30 min after exposure. More plots of this animal, including the entire time series, can be found in Kvadsheim et al. (2015).

Based upon these plots, behavioural changes were described and scored by two independent groups of experts in accordance with the Southall et al. (2007) severity scale modified by Miller et al. (2012) (Table 1). In the present study, we made some additional modifications based upon observed behavioural changes that were not covered in the existing scale, mainly changes with shorter duration (Table 1).

The severity of a behavioural response depends upon its duration, but the duration of the response will likely vary with exposure duration. Experimental exposures typically are shorter than real naval sonar operations; and to enable extrapolation, the Southall et al. (2007) severity scale defined duration on a time scale relative to exposure duration. A *moderate* change in behaviour had duration similar to the duration of the exposure, and *minor* or *brief* responses were considered to be progressively shorter than the exposure; whereas a *prolonged* response was significantly longer than the exposure, considering the behavioural and diving patterns of the whale species. For some response types, a *long-term* or *extended* response is also included to account for responses that are exceptionally long (i.e., longer than prolonged) (Table 1). This logic was used in previous studies using the severity scale (Miller et al., 2012) and is also used here; however, the humpback whale exposures were very short (5 to 10 min) compared to exposures to other species. To make the results more comparable across species, for humpback whale data, a behavioural change lasting approximately 10 min (same as duration of the exposure) to 1 h was defined as *moderate*, while *prolonged* was used for responses lasting for more than 1 h. Moreover, a *brief* change was defined as significantly shorter than the exposure (less than 5 min), while *minor* was used for changes of duration between *brief* and *moderate* (5 to 10 min). Using these criteria for the humpback whale data enabled comparison with the other two species of this study as well as the three species studied by Miller et al. (2012), which were all exposed for a much longer period (30 to 60 min) than the humpback whales.

One scoring group consisted of authors PM, SI, CC, and PW, and the second group of authors PK, FL, LS, LK, and FV. Both groups conducted an initial independent scoring that was blind to the scoring of the other group but not blind to the experimental condition. The results of the two groups were tabulated before they met with an adjudicator (author PT) to reach a consensus scoring. The judgment of whether behavioural changes were scored as responses took into consideration the variation in this behaviour during the pre-exposure baseline period. For example,

if a change in behaviour during exposure also happened regularly during baseline, this change would generally not be considered as a response but a coincidental change in behaviour. In some cases, this was difficult to judge; and in these cases, behavioural changes were scored as responses with *low confidence*. If a behavioural change during an exposure was identified but was not considered to be a response, this change was given a 0 score (no response).

The data were then inspected in detail to determine the precise onset time of each behavioural change as well as its duration. For sonar exposures, the highest SPL received up to the onset of the behavioural change (SPL_{max}) was used as the received level associated with the scored response. We also report the received SEL_{cum} until the last ping before the response (Appendix A, Tables AI through AIII).

Statistical Analysis

Statistical analysis was only carried out for humpback whales as the sample sizes for the minke and bottlenose whale ($n = 1$ for both) were too low for statistical comparison. The statistical analysis aimed to test the null hypothesis that the number and severity of scored responses were random with respect to the type of exposure (sonar – no-sonar-control – killer whale playback – broadband noise playbacks).

Generalized estimating equation (GEE) models were fitted to the data to test the null hypothesis while accounting for serial correlation within each individual and to obtain effect estimates averaged across individuals. GEEs are a means to estimate the parameters of a regression model with unknown correlation between outcomes. All models were fitted with the “genmod” procedure in SAS 9.3 (Littell et al., 2006).

Severity scores were expressed as two independent response variables: (1) number of scored responses ($\#Score$) per exposure session, modelled as a Poisson variable (identity-link); and (2) maximum severity score ($maxScore$) per exposure session, modelled as a multinomial ordinal variable (cumulative logit link). The *no-sonar control* was selected as the intercept within the model, and candidate explanatory variables were designed to capture differences from this baseline. Six explanatory variables (covariates) were defined present (1) or absent (0). Covariate S was set present during all sonar exposures, $kwPB$ during killer whale sound playbacks, and *no-ramp-up* during sonar exposures that did not include a ramp-up period. To test for sensitization/habituation to sonar during the second sonar exposure compared to first, $S2$ was set present during the second sonar exposure. The effect of the killer whale playback

Table 1. Severity scale for scored behavioural responses used in this study. The scale is based on the original scale of Southall et al. (2007). Behavioural changes in **bold** are modifications made by Miller et al. (2012) and in **bold italics** are modifications made in the present study. Brief responses – duration significantly shorter than the exposures; Minor responses – shorter than the exposures, but longer than Brief; Moderate responses – roughly the duration of the exposure; and Prolonged responses – significantly longer than the exposures. Long-term and Extended responses are both considered to be exceptionally long duration responses (longer than Prolonged).

Score	Behavioural responses
0	No observable response
1	Brief orientation response
2	• Moderate or multiple orientation responses • Brief or minor changes in respiration rates • Brief cessation/modification of vocal behaviour • <i>Brief change in dive profile</i>
3	• Prolonged orientation behaviour • Minor change in locomotion (speed/direction) and or dive profile but no avoidance of sound source • Minor cessation/modification of vocal behaviour • Individual alert behaviour • Moderate change of respiration rate
4	• Moderate change in locomotion (speed/direction) and or dive profile but no avoidance of sound source • Brief avoidance of sound source • Minor shift in group distribution • Moderate cessation/modification of vocal behaviour • <i>Brief cessation of feeding</i>
5	• Extended change in locomotion (speed/direction) and or dive profile but no avoidance of sound source • Minor avoidance of sound source • Moderate shift in group distribution • Change in inter-animal distance and/or group size • Prolonged cessation or modification of vocal behaviour • <i>Minor cessation of feeding</i>
6	• Moderate avoidance of sound source • Extended cessation or modification of vocal behaviour • Visible startle response • Moderate cessation of feeding • Prolonged shift in group distribution • Brief or minor separation of female and dependent offspring • Aggressive behaviour related to noise exposure • Brief cessation of reproductive behaviour • Moderate cessation of resting behaviour
7	• Prolonged cessation of feeding • Moderate separation of female and dependent offspring • Severe and or sustained avoidance of sound source • Extensive or prolonged aggressive behaviour • Clear anti-predator response • Moderate cessation of reproductive behaviour • <i>Prolonged avoidance</i>
8	• Obvious aversion and/or progressive sensitization • Long-term avoidance of area • Prolonged or significant separation of female and dependent offspring with disruption of acoustic reunion mechanisms • Prolonged cessation of reproductive behaviour
9	• Outright panic, flight, stampede, or attack • Avoidance related to predator detection

boat was modelled by setting *PB* present for both killer whale sounds and noise playbacks. The order of the playbacks was modelled by setting *PB2* present during the second playback.

To test for effects of the sonar and killer whale sound playbacks, model selection was carried out on the explanatory variables (*S*, *S2*, *PB*, *PB2*, *kwPB*, and *no-ramp-up*) fitted to scoring data from the humpback whale experiments. Backwards step-wise selection was conducted based upon Type 3 Wald tests for each explanatory variable. Variables were removed from the full model one at a time until *p* values were smaller than 0.05 for all covariates. The routine was carried out separately for #Score and maxScore.

In addition, dose response severity functions were generated for humpback whales using recurrent event survival analysis wherein marginal stratified Cox proportional hazards models were fitted to the severity score data (see Harris et al., 2015, for full details of model application to severity score data). In short, these models allow us to combine the results from individual exposures to estimate the likelihood of response as a function of exposure level (dose) and behavioral or contextual covariates. We can also stratify by response severity to produce functions for different severity levels. Recurrent event survival analysis is generally used to assess time-to-event, particularly in the medical context; however, we were interested in acoustic dose-to-event, and so we replaced time with received acoustic energy in the form of SEL_{cum} . These models, therefore, provide a relationship between dose SEL_{cum} and the probability of response of different severity levels. Similar analysis has been carried out for pilot, killer, and sperm whales (Harris et al., 2015, based on Miller et al., 2012); therefore, we can compare responsiveness between these four species by extracting the 50% probability of response from the dose-response functions for each species.

Our main interest is behavioural changes with the potential to affect vital rates (i.e., severity score 4 to 6). For humpback whales, we identified the first occurrence of each response level (severity scores 1 to 3, 4 to 6, and 7 to 9) within each exposure session for inclusion in the model. In the case of no response across all levels within an exposure session, we allocated the SEL_{cum} at the end of the exposure session, and the data were labeled as right-censored. We fitted models to data from the sonar exposure experiments and included exposure history (1st or 2nd exposure sessions) as a continuous covariate and behavioural state as a factor covariate.

Behavioural state—feeding or non-feeding—was determined by the presence of lunges in the dive record 5 to 10 min (depending on the exposure

duration) prior to exposure, and feeding was assigned if any lunging behaviour was observed in this pre-exposure period. The data were collected over 20 exposure sessions across 10 different individuals. We fitted all possible model combinations and used an Akaike Information Criterion (AIC)-based model selection. For the selected model, we tested that the proportional hazards assumption was met (Kleinbaum & Klein, 2005; Harris et al., 2015). These analyses were carried out in R, Version 3.0.2 (R Core Team, 2013) using the survival package (Therneau, 2014).

Results

We conducted a total of 13 controlled sonar exposure experiments with humpback ($n = 11$), minke ($n = 1$), and bottlenose whales ($n = 1$). Multiple exposure sessions were conducted with a subset of the tagged animals, giving a total of 12 no-sonar control approaches, 22 sonar exposure sessions, and 8 killer whale and noise playbacks (Table 2).

Summary of Behavioural Responses

A total of 66 changes in behaviour were judged (scored) to be responses to the experimental condition in the 50 exposures conducted (Tables AI through AIII). The number of changes in behaviour scored as responses ranged from zero change in 17 exposure sessions to six changes in one exposure session (ha13_176a). Zero, one, or two changes were most common per exposure session (17, 15, and 14 exposure sessions, respectively). When more than one change in behaviour was scored in the same exposure, these scored changes often tended to reflect different aspects of a change in behavioural state (e.g., cessation of feeding was often associated with a simultaneous change in dive pattern (e.g., mn12_180 sonar1; see Figure 1).

Overall, the most common response scored in all three species to all exposures was horizontal avoidance (Figure 4) of the sound source or from its projected future path, including one case of long-term avoidance of the area (bottlenose whale), and one case of progressive aversion (minke whale) in which the whale progressively increased speed and changed dive pattern in order to more efficiently move away from the sound source. Movements that were not aimed at increasing the distance to the source were not scored as avoidance responses but, rather, as changes in orientation or locomotion.

The second most commonly observed change in behaviour was a change in the dive profile. Many of these changes in diving pattern were also associated with cessation of feeding (e.g., mn12_170 killer whale playback; Kvadsheim

Table 2. List of the experiments for the three species. Tag id is the code for the tag deployment, and the number of exposures of each type is given. Details of all deployments can be found in Kvadsheim et al. (2015).

Species	Year	Tag id	No-sonar control	Sonar	kwPB	Noise PB	Comments
Humpback whale (<i>Megaptera novaeangliae</i>)	2011	mn11_157a	1	2	1	1	
	2011	mn11_158ab	1				Two tags on the same animal
	2011	mn11_160ab	1	2	1	1	Two tags on the same animal
	2011	mn11_165def	1	2	1	1	Three tags on two different animals
	2012	mn12_161ab	1	2	1	1	Two tags on the same animal
	2012	mn12_164ab	1	2	1	1	Tags on two different animals
	2012	mn12_170ab	1	2	1	1	Two tags on two different animals
	2012	mn12_171ab	1	2	1	1	Two tags on the same animal
	2012	mn12_178a	1	2			
	2012	mn12_179a	1	2			
	2012	mn12_180ab	1	2	1	1	Two tags on the same animal
Minke whale (<i>Balaenoptera acutorostrata</i>)	2011	ba11_180a	1	1			Ctag without acoustic sensors
Bottlenose whale (<i>Hyperoodon ampullatus</i>)	2013	ha13_176a		1			Lost visual contact with animal prior to exposure

et al., 2015). Cessation of feeding was not scored for the minke whale because the identification of lunges was based on the acoustic sensor, which was lacking in the tag used in this particular trial. Changes in group distribution were less common (Table 2) and only scored for humpback whales. However, humpback and minke whales are often solitary in these waters; indeed, five of 11 humpback whales and the minke whale were solitary for most of the period they were tracked. Changes in vocal behaviour were only scored for the bottlenose whale. The bottlenose whale (ha13_179a) stopped producing foraging sounds completely during sonar exposure (Figures 3 & 4). The two most severe responses scored were observed in response to sonar: long-term avoidance of a larger area (score 8) for the bottlenose whale and obvious progression of the avoidance response (score 8) for the minke whale.

Severity of Scored Responses in Relation to Received Sound Pressure Level

For humpback whales, changes in behaviour were scored with response thresholds over a broad range of received sound pressure levels (SPL_{max}) from 94 to 179 dB re 1 μ Pa (Table 3; Figure 5). The highest score (severity 7) corresponded to an onset level of 164 dB re 1 μ Pa. The majority of scores (16 of 27) were at threshold levels above 140 dB re 1 μ Pa.

Neither of the covariates investigated in the Cox proportional hazards model were significant at the $p = 0.05$ level and, therefore, a model with no covariates was fitted to produce dose-response functions across all humpback whales for responses of different severities. From the function for responses with potential to affect vital rates (severity score 4 through 6) (Figure 6), we can see that the probability of a response increased as received SEL_{cum} increased, with 50% probability at 179 dB re 1 μ Pa²s (CI 141 to > 185 dB re 1 μ Pa²s, $n = 20$). Given that behavioral state was not a significant covariate for humpback

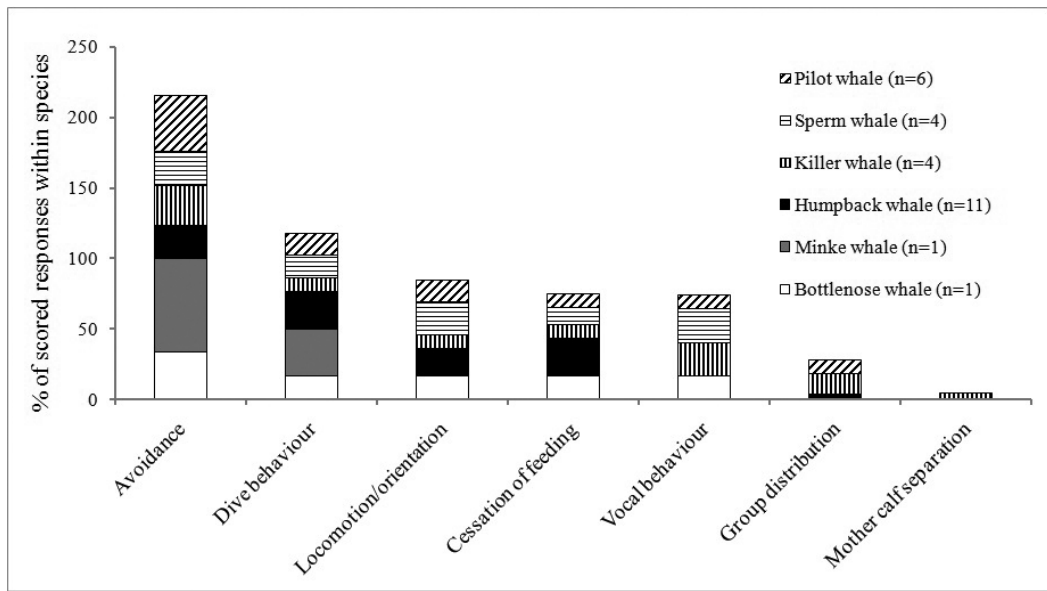


Figure 4. Percentage of observed behavioural response types within each species in response to sonar. The figure includes data for six species: the three species of the current study and the three species studied by Miller et al. (2012). The sonar signals were presented to the animals in a random behavioural context. This implies that some type of responses were not available to some individuals at the time of the exposure—for example, only feeding animals can cease feeding, and only vocalizing animals can stop vocalizing. However, since real sonar exercise also starts at a random time relative to the animal's behavioural context, this figure represents an expected distribution of response types during real sonar exercises.

whales, we can only compare humpback whales averaged across all behavioral states with feeding and non-feeding killer, pilot, and sperm whales (from Harris et al., 2015). For feeding killer, pilot, and sperm whales exposed to low-frequency active sonar (LFAS), the 50% probability of response with potential to affect vital rates was 89 dB re 1 $\mu\text{Pa}^2\text{s}$ (CI = 80 to > 180 dB re 1 $\mu\text{Pa}^2\text{s}$, $n = 1$), 142 dB re 1 $\mu\text{Pa}^2\text{s}$ (CI = 126 to > 180 dB re 1 $\mu\text{Pa}^2\text{s}$, $n = 1$), and 133 dB re 1 $\mu\text{Pa}^2\text{s}$ (CI = 122 to > 180 dB re 1 $\mu\text{Pa}^2\text{s}$, $n = 4$), respectively. For non-feeding killer whales, the 50% probability of response at the same severity level was 165 dB re 1 $\mu\text{Pa}^2\text{s}$ (CI = 157 to > 180 dB re 1 $\mu\text{Pa}^2\text{s}$, $n = 2$). Non-feeding pilot whales did not reach the 50% probability of response at the maximum exposure level 179 dB re 1 $\mu\text{Pa}^2\text{s}$, and sperm whales were always feeding at the start of the exposure. In all comparisons, the humpback whales had a higher SEL_{cum} at the 50% response probability than the other three species,

Two of three of the scored responses for the minke whale had thresholds ≤ 146 dB re 1 μPa ; and in the bottlenose whale, four of six responses happened at thresholds ≤ 130 dB re 1 μPa (Table 3; Tables AI–AIII). The lowest threshold in these two species was at 83 and 86 dB re 1 μPa in the minke whale and bottlenose whale, respectively, but these

responses were scored to have modest severities of only 2 and 3, respectively. The highest severity scored in minke whales and bottlenose whales was severity 8 for both species, with thresholds of 158 and 130 dB re 1 μPa , respectively.

Severity of Scored Responses in Relation to Exposure Type

For humpback whales, the scored changes in behaviour were generally more severe for the killer whale playback and sonar exposures than for the no-sonar control and noise playbacks (Table 4). Sonar exposure and killer whale sound playbacks were estimated to have 8 and 40 times the odds of a higher maximum score than the no-sonar control and noise playbacks, respectively (odds ratio; Table 5). The average number of scores was estimated to be over 1.35 times higher during sonar exposures and 2.25 times higher during killer whale sound playbacks, compared to no-sonar control transmission for humpbacks (Table 5). Thus, signal type (kwPB and S) explained a significant amount of variation in the severity scores across exposure types, both in terms of maximum severity score and number of scores per exposure (Table B1: Type III Wald test, $\chi^2 = 5.4$ to 31.5, $p < 0.021$). Maximum severity and number of scores were in general higher for the first sonar exposure

Examples of Severity Scoring

The tables in Appendix A show the complete set of scored responses. The complete set of data plots from all scored sessions can be found in Kvadsheim et al. (2015). Three examples of data plots used as a basis for expert scoring are shown in Figures 1 to 3, one for each species. The scoring results of these experiments can be seen in the following examples:

Example 1. Humpback Whale (mn12_180ab), Sonar 1 – Three different behavioural responses were scored for this exposure. This experiment was conducted on a mother-calf pair, with the mother as the tagged animal. We did not manage to tag the calf, but the behaviour of the two whales at the surface was often synchronous. The data plots for this exposure are shown in Figure 1.

Severity	Description
5	<i>Minor avoidance</i> – After a period of tortuous horizontal movement prior to the exposure, the track became more directional while the animal was moving out of the path of the ship. The avoidance response started almost immediately after the first pulse was transmitted but did not last beyond the duration of the sonar exposure.
4	<i>Moderate change in dive profile</i> – The animal made one deep dive to 100 m that started 2.5 min after sonar onset. A dive this deep was unusual and did not happen in the pre-exposure tag record. The animal's dive behaviour changed after this deep dive. Normal diving resumed within 1 h after exposure. The start of the deep dive was considered to be the onset of the response.
6	<i>Moderate cessation of feeding</i> – The animal conducted shallow feeding dives to 10 to 40 m depth with frequent lunges almost continuously for a period of 7 h before the sonar exposure. A few minutes into the exposure, the animal made one last feeding dive with one lunge before it changed dive behaviour and stopped lunging. Some feeding resumed within 1 h. The start of the unusual deep dive that followed the dive with a lunge was taken as the onset of the response.

Example 2. Minke Whale (ba11_180a), Sonar 1 – This was the only minke whale tagged. This whale was subject to a single sonar exposure session after a no-sonar control approach. Three different behavioural responses were scored for the sonar exposure. The data plots for this exposure are shown in Figure 2.

Severity	Description
2	<i>Brief change in dive profile</i> – The animal changed its dive pattern at sonar onset from very shallow diving (0 to 5 m) to diving into the sound channel at around 50 m depth. This started at the first sonar ping at a received SPL of 83 dB re 1 μ Pa. During the 30 min prior to sonar onset, the whale had been doing only shallow (< 10 m deep) dives. This behaviour could be a possible attempt to improve the localization of the sound source.
7	<i>Prolonged avoidance</i> – The animal increased speed and swam directly away from the sound source throughout the rest of the exposure. Opportunistic visual observations of skim feeding at the surface before the start of the sonar exposure indicated that this response might also have involved a cessation of feeding. The start of the response was taken as the time of the first sighting where the animal had increased speed and swam at a course directly away from the source at a received SPL of 146 dB re 1 μ Pa.
8	<i>Obvious progressive aversion (and sensitization)</i> – The animal continued to increase its speed as the exposure progressed, swimming at such a high speed that the distance to the source ship remained constant. About halfway through the exposure, the dive pattern changed to shallower diving, which may be a way to move more effectively away from the source. This shift in dive pattern is set as the onset of this response with a received SPL of 158 dB re 1 μ Pa.

Example 3. *Bottlenose Whale (ha13_179)*, *Sonar 1* – This was the only bottlenose whale that was tagged, and it was subject to a single sonar exposure experiment. Six different behavioural responses were scored. The data plots for this exposure are shown in Figure 3.

Severity	Description
3	<i>Minor change in locomotion</i> – The whale swam towards the source at the very beginning of the exposure before turning away later. Onset was set to when the animal turned towards the source.
5	<i>Extended change in dive profile</i> – The whale conducted a very deep and unusual dive during the sonar exposure. This dive seemed to start as a shallow dive, but just after the ascent had started, the animal dove again, making the deepest dive ever recorded for a bottlenose whale down to 2,340 m. After this deep dive, the animal made shallow dives to about 250 m depth throughout the rest of the tag record without performing any deep dives. The onset of the response was taken as the start of the descent of the deep dive.
7	<i>Prolonged cessation of feeding</i> – The whale conducted regular foraging dives to 1,500 to 2,000 m depth in the 5 h prior to exposure, with resting dives to about 100 to 200 m between the deep dives. After sonar onset, no feeding activity (clicks or buzzes) was recorded for the rest of the record. The onset of the response was chosen as the time during the deep dive when the animal reached the average depth at which feeding sounds had started in the previous deep dives (168 m). The animal had still not returned to feeding behaviour at the end of the tag record (7 h later), and the response, therefore, was scored as prolonged.
7	<i>Prolonged avoidance</i> – After a small loop towards the source scored as a minor change in locomotion, the animal turned away and moved consistently away from the source with a constant course. The animal was moving away from the area throughout the tag record, and the response, therefore, was scored as prolonged. The onset was taken as the point when the animal abruptly changed course and started moving away from the source.
8	<i>Long-term avoidance of area</i> – Visual and acoustical detections of bottlenose whales in a surveyed 14 × 18 km area around the sonar exposure dropped dramatically (Kvadsheim et al., 2014; Miller et al., 2015). In the 6-h period after sonar exposure, no acoustic detections and very few visual detections were made compared to the 24-h period before sonar. Additionally, the tagged whale avoided the area for at least 7 h after the exposure. As no information was available on the full duration of avoidance of the experiment area, we scored the response as greater than prolonged following precautionary interpretation of the data. The onset of this response was taken to be the same as the onset of the prolonged avoidance of the tagged animal.
6	<i>Extended cessation of vocal behaviour</i> – The animal was vocally active during entire tag record prior to sonar exposure but turned completely silent for more than 7 h after until the tag detached and the record ended. Onset of the response was the same as for cessation of feeding.

compared to the second sonar exposure (Tables 3 & 4). However, covariate S2 was the last covariate to be dropped in the backwards step-wise model using Type III Wald tests for both number of scores and maximum score, indicating a trend in the data that the second sonar exposure elicited less severe and fewer responses than the first exposure, although these order effects were still not quite significant ($p = 0.09$ and $p = 0.13$ for #Score and maxScore, respectively).

Additionally, the minke whale had much higher severity responses during the sonar exposure than the no-sonar control exposure, although this could not be tested statistically.

Potential False Positives

Of all identified behavioural changes during experimental conditions, 28% were scored as severity 0 (no response). In five exposures, a change in behaviour was scored as a response but with low confidence whether this change in behaviour was a response or not (Table 3). Details of these responses are found in Table A1.

Discussion

The objectives of the present study were to identify behavioural responses of cetaceans to sonar

Table 3. Scored responses during sonar exposure with associated maximum received sound pressure level (SPL_{max}) in 10 dB bins for the three study species. “1” indicates a score given during the first sonar exposure (sonar 1), while “2” indicates scores for the second sonar exposure (sonar 2) of the same animal. Scores in parentheses indicate *low confidence* scores.

Experiments with humpback whales												
SPL _{max} (dB re 1 μPa)												
Severity score	70-79	80-89	90-99	100-109	110-119	120-129	130-139	140-149	150-159	160-169	170-179	180-189
9												
8												
7												
6												
5												
4												
3												
2												
1												

Experiment with minke whale												
SPL _{max} (dB re 1 μPa)												
Severity score	70-79	80-89	90-99	100-109	110-119	120-129	130-139	140-149	150-159	160-169	170-179	180-189
9												
8												
7												
6												
5												
4												
3												
2												
1												

Experiment with bottlenose whale												
SPL _{max} (dB re 1 μPa)												
Severity score	70-79	80-89	90-99	100-109	110-119	120-129	130-139	140-149	150-159	160-169	170-179	180-189
9												
8												
7												
6												
5												
4												
3												
2												
1												

Table 4. Number of exposure sessions sorted by maximum score within each exposure and stimulus type. Numbers in parentheses are minke whale data, numbers in brackets are bottlenose whale data, and numbers not in brackets/parentheses are humpback whale data. Numbers in **bold** (bottom line) combine data from all three species. Total number of responses includes all scored responses—not just the maximum score. The bottom row shows percent of the sessions that have a score of 4 or higher, thus, having the potential of affecting vital rates.

Severity	No-sonar control	Sonar 1	Sonar 2	Killer whale PB	Noise PB
9					
8		(1) [1]			
7		1		4	
6	1	3	2	2	1
5		1	1	1	1
4	1	2	1		1
3	2		2		
2	-1	2	1	1	
1					
0	8	2	3		5
Total # of exposures	13	13	10	8	8
Total # of responses	7	24	9	18	9
% score ≥ 4	15.4	69.2	40	87.5	37.5

Table 5. *Generalized Estimating Equation (GEE) model estimates:* For number of scores (#Score), the estimates are the number of maximum scores averaged across individuals in each exposure condition (intercept: no-sonar control and noise playback; S: sonar approach; and kwPB: playback of killer whale sounds). For maximum score (maxScore), the probability of each maximum score is shown for the intercept condition (no-sonar control and when the noise playback was presented first). The odds ratios give the odds of a higher severity score in the presence of each exposure compared to the odds during the no-sonar control (intercept). The odds ratio for a higher maxScore is shown in **bold** for each exposure condition. S is all sonar exposures, kwPB is killer whale playbacks, and PB2 is the second playback. Test statistics show Z score and *p* value for the null hypothesis that the respective model estimates (intercepts and coefficients for each explanatory variable) equal zero.

#Score	Estimate	95% CI		Z score	<i>p</i> value
Intercept	0.4211	0.1775	0.6646	3.39	0.0007
S	1.35	0.3209	2.3791	2.32	0.0205
kwPB	2.25	1.3674	3.1326	5.61	0.0001
maxScore	Estimate	95% CI		Z score	<i>p</i> value
7	0.01	0	0.03	-5.94	0
6	0.04	0.01	0.15	-4.59	0
5	0.03	0.03	0.22	-3.99	0
4	0.05	0.05	0.29	-3.6	0
3	0.06	0.07	0.41	-2.55	0.011
2	0.08	0.12	0.49	-2.08	0.037
0	0.73	0.51	0.88	—	—
S	8.28	1.69	40.52	2.61	0.009
PB2	12.52	1.59	98.54	2.4	0.016
kwPB	40	5.58	286.78	3.67	0

and control signals by use of expert judgment, to describe the severity of the responses using standardized criteria (Southall et al., 2007; Miller et al., 2012), and, finally, to define the received level at which these scored responses began.

Experimental Design and Scoring Methodology

Severity of the observed responses ranged from unlikely to likely to affect vital rates, with onset levels ranging over a wide range of received sound pressure levels (SPL_{max}) and sound exposure levels (SEL_{max}). Due to the short exposure duration of our experiments, however, it is unlikely that the experimental animals themselves actually suffered any effect on their vital rates. Still, higher severity responses to sonar in real-world scenarios in which the exposures could be significantly longer or repeated could potentially affect the animals in a biologically significant way.

The terms *brief*, *minor*, and *moderate* used in the severity scale to characterize a change in behaviour have been defined differently between studies. Williams et al. (2014) defined these terms with respect to a percent change in a variable compared to baseline—a method that can be relatively easily applied when evaluating changes in numeric variables such as change in speed or directedness. However, not all of our parameters could be quantified in this manner (e.g., feeding activity, vocalization[s], and social behaviour). Southall et al. (2007) used the duration of the change in behaviour relative to the duration of the exposure to quantify change(s). Miller et al. (2012) used this same approach.

Herein, we also applied this same method, thus considering duration of response relative to duration of exposure. A full-scale naval exercise may last for hours, or even days, and a behavioural response such as cessation of feeding, which lasts the full duration of this exposure, could have severe consequences during the feeding season. In our results, moderate or prolonged responses continued beyond the end of our relatively short duration experimental sonar exposures sessions. The extent to which extrapolation can be made from short duration experimental exposures to long duration naval scenarios using the severity scale has not been tested.

A goal of the experimental design was to reproduce a realistic exposure of naval sonar, with source levels, frequency bands, and waveforms common to a number of nations' sonar operations (Ainslie, 2010). Some aspects of our experimental design also represented fairly realistic sonar operation scenarios with a moving source transmitting sonar signals at high source levels with a regular ping interval. Our sonar exposures also started randomly with respect to the context or behavioural state of the tagged animal, which is also the case

for actual naval sonar operations. The approach towards the animal by the source ship was necessary to achieve an escalating received level of the sonar pulses on the animal. However, this brought the sonar source quite close to the animal, and this might not be very realistic for operational sonar use except in rare circumstances.

Behavioural changes that did not commonly occur during the baseline (pre-exposure) period were scored as responses to the experimental stimuli. In some cases, behavioural changes might have been caused by unknown factors and not by responses to the experimental stimuli—for example, in mn11_165 sonar 1 (Kvadsheim et al., 2015), no lunges were made by the animal for a period of three dives (8 min) during exposure. However, similar periods without lunges were common in the baseline phase as well as throughout the rest of the tag record, and, therefore, this behavioural change was not scored as a response to the exposure (severity 0). A more difficult example is mn12_170ab sonar 1 (Kvadsheim et al., 2015). Two animals were travelling together; both were tagged. One animal started feeding 6 min before exposure, after a long period (> 7 h) of not feeding, but stopped feeding during exposure. Such short duration feeding events were very rarely seen in the humpback whale records which indicates a response. However, the other animal travelling alongside did not feed at all in the periods immediately before, during, or after exposure. This might indicate a low-quality food patch, for example, and the first animal may simply have realized this and stopped feeding (i.e., no response). As a precautionary approach, we scored this as cessation of feeding (severity 7 based on the duration), but we gave it a low confidence score.

Patterns of Severity by Received Level and Species Differences

The single exposures with minke and bottlenose whales made it challenging to compare between the three species. A range of responses and severities were scored at a very broad range of received levels for all species (Table 3; Figure 5). However, the single sonar exposures with minke and bottlenose whales showed that both these animals responded with highly severe changes in behaviour (severity 7 & 8), and the threshold of these responses were also relatively low: SPL_{max} = 146 to 158 dB re 1 μ Pa in minke whales and 130 to 141 dB re 1 μ Pa in bottlenose whales. Compared to this, responses to sonar by humpback whales were generally less severe and were triggered at higher received levels. The maximum severity in humpback whales was a prolonged cessation of feeding, scored at a severity 7 with a response threshold at SPL_{max} = 164 dB re 1 μ Pa. This single

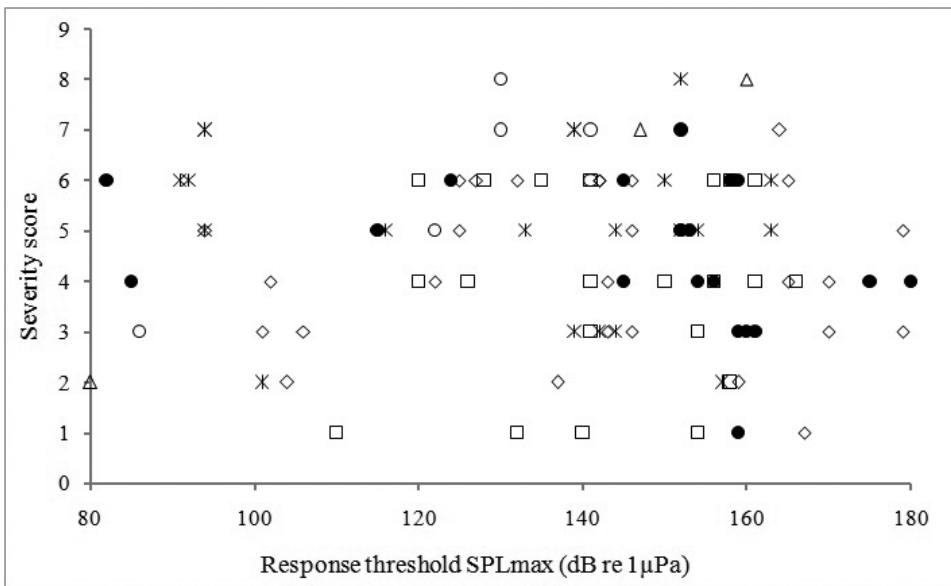


Figure 5. Scored severity vs threshold of all responses in the three species of this study and the three species of Miller et al. (2012): humpback whale (◇), minke whale (△), bottlenose whale (○), pilot whale (●), killer whale (*), and sperm whale (□).

event was also a response judged to be of low confidence that it actually was a response to the sonar and not simply an incidental change in behaviour (Table 3; Table AI). Thus, with the reservation inherent in drawing conclusions from single experiments, our results suggest that minke and bottlenose whales are more sensitive to sonar than humpback whales. The additional observation was that the number of bottlenose whales feeding in the larger exposed area in the 6 h following the sonar transmission dropped dramatically (Miller et al., 2015). This may indicate that, at least for bottlenose whales, the observed response was not limited to just the single tagged animal.

Other studies that examined humpback whale reactions to sonar have shown weak avoidance (Maybaum, 1993) and changed singing activity (Miller et al., 2000) but with large individual variation. The number of acoustic detections of minke whales was found to drop markedly during naval sonar activity (Martin et al., 2015). No other studies have been conducted with bottlenose whales, but other beaked whale species have shown strong avoidance of sonar signals (Tyack et al., 2011; DeRuiter et al., 2013; Moretti et al., 2014), and both minke and beaked whales have been involved in strandings that were linked to naval sonar (Balcomb & Claridge, 2001; D'Amico et al., 2009), which is not the case for humpback whales. These studies are in accordance with the present results indicating all three species do respond to 1 to 2 kHz sonar signals and that minke

and bottlenose whales appear to be more sensitive to this disturbance than humpbacks.

Our results further imply that the severity of responses cannot be accurately predicted from the received levels alone (Table 3; Figure 5) because responses with potential to lead to changes in vital rates (> 4) were occasionally triggered at very low levels (< 100 dB re 1 μ Pa) (Table 3; Figure 5). This lack of correlation between severity and received levels was also noted by Miller et al. (2012), who concluded that all exposures above hearing threshold implied some risk of triggering a relatively severe behavioural response. Thus, other elements, such as the context in which a whale is exposed to sound, also influence its response.

We used marginal Cox proportional hazards models to produce dose-response functions for behavioral responses of different severities for the humpback whales to allow comparison with previous analysis on killer, sperm, and pilot whales (Harris et al., 2015). We limited covariates considered herein to those considered for other species, but neither behavioral state nor exposure order significantly contributed to the fit of the model to humpback whale data. We believe there may be other covariates that could be considered with the data on humpback whales, but that was beyond the scope of this study.

The comparison between all four species indicated that, on average, humpback whales respond at higher doses of accumulated sound than pilot, killer, and sperm whales, indicating humpback

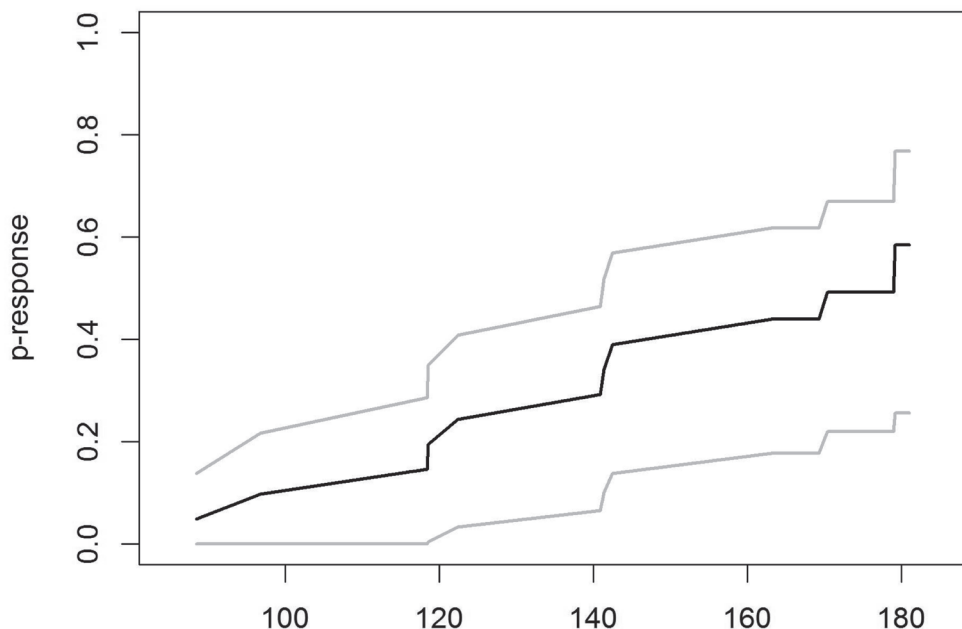


Figure 6. The predicted probability of a response of severity with potential to affect vital rates (severity 4 to 6) vs received cumulative sound exposure (SEL_{cum}) (dB re $1 \mu Pa^2 s$) for humpback whales; the grey lines represent the 95% confidence intervals.

whales to be less sensitive than the other three species. However, the sample size for some of these species is rather low in some cases, and, therefore, these comparisons should not be over-interpreted.

Although not statistically significant, exposure history (1st or 2nd exposure session) was close to significant ($p = 0.07$) for humpback whales, with a lower probability of response during the second exposure compared to the first. This indicates that there might be some habituation between the first and second sonar exposure. Differences in responsiveness to sound levels may also be due to differences in sensation levels between species; however, studies of pilot whales (Antunes et al., 2014) and killer whales (Miller et al., 2014) indicated no striking difference in response threshold between 1 to 2 kHz and 6 to 7 kHz sonar, despite a 30 to 50 dB difference in hearing sensitivity of killer whales (Szymanski et al., 1999; Miller et al., 2014).

For the three species in the current study, the hearing thresholds have also not been well examined. No direct measurement of hearing ability has been made for any baleen whale, but biomechanical modelling of the hearing organ of baleen whales indicates that minke whales and humpback whales both have sensitive hearing in the 1 to 2 kHz band tested, although absolute hearing thresholds were not provided (Houser et al., 2001;

Tubelli et al., 2012; Cranford & Krysl, 2015). The hearing ability of bottlenose whales has not been measured, but beaked whale species have their greatest hearing sensitivity at around 40 kHz (Cook et al., 2006; Finneran et al., 2009), although hearing has not been tested below 5 kHz. The lack of quantitative information on hearing thresholds in the relevant frequency band for all our study species makes any attempt to explain differences in responsiveness with hearing ability (sensation level) speculative.

Overall, avoidance of the sound source was the most commonly scored response to sonar, both in the three species studied here (humpback, minke, and bottlenose whales) and in the three species studied by Miller et al. (2012) (pilot, killer, and sperm whales) (see Figure 4). Changes in dive behaviour and other changes in movement and orientation were also commonly seen in all these species. Change in vocal behaviour was the second most common type of response observed in all four odontocetes, which were all very vocal in the baseline period. Cessation of feeding was also recorded in all species except for minke whale, for which the tag did not contain sensors that could detect feeding behaviour.

Patterns of Severity by Exposure Type in Humpback Whales

For humpback whales, there were large differences both in the number of scored responses and the maximum severity of responses scored within an exposure depending upon exposure type (e.g., sonar, no-sonar control, killer whale, or noise playback) (Table 5; Table B1). Killer whale playbacks had a higher likelihood of being assigned both a higher number of scores per exposure and scores of higher severity than sonar exposure (Tables 4 & 5; Table B1). Thus, humpback whales reacted less strongly to exposure to naval sonar than to the playback of a natural predator. Killer whales can potentially attack all three target species (Jefferson et al., 1991); and killer whale sounds were expected to induce a natural anti-predator response. Our intention was to conduct killer whale playback experiments with all three target species of our study, but minke whales and bottlenose whales responded so strongly to the sonar exposure, which was conducted first, that exposure to killer whale sounds was not completed with those individuals.

Overall, humpback whales responded consistently and strongly to killer whale sound playbacks, whereas they rarely changed behaviour in response to the broadband noise playback, thus, specifically reacting to the detection of killer whale sounds (Curé et al., 2015). Killer whale sound playbacks induced horizontal avoidance responses away from the sound source and changes in the dive profile as well as cessation of feeding for whales that were foraging before the start of exposure. Among species that can potentially be preyed upon by the killer whale, humpback whales are thought to be regular targets (Jefferson et al., 1991; McCordic et al., 2013). In the present study, responses of humpback whales to killer whale playbacks may be interpreted as typical anti-predator behaviours (Frid & Dill, 2002) wherein animals stop fitness enhancing activities such as feeding when facing a perceived risk of predation (Curé et al., 2015).

Sonar exposure was the second strongest stimulus causing behavioural responses in humpback whales (Tables 4 & 5). Although the statistical analysis did not fully support it, there did appear to be a trend in the data suggesting that for humpback whales, the second sonar exposure resulted in less severe and fewer responses (Tables 3 & 4; Table B1). Consistent with the findings of the marginal Cox proportional hazards model reported above, this difference in the first and second sonar exposure sessions may suggest habituation to the sonar in that the humpback whales responded less intensely to the familiar sound that had not led to acute danger when first heard.

Conclusions

Based upon detailed expert examination of data recorded during experimental exposures to sonar, we found that exposure to 1 to 2 kHz naval sonar signals changed the behaviour of whales. The descriptive approach taken in this study is useful to present the data-rich outcomes of experimental exposure of sonar to cetaceans, which can be difficult to analyze statistically. Even though comparison of the three species studied here was limited to the interpretation of single exposures to sonar for minke and bottlenose whales, those two experiments indicate high sensitivity to disturbance from sonar sounds. The bottlenose whale examined showed behavioural responses with high severity at a relatively low response threshold. This is consistent with studies of other species of beaked whales. The minke whale examined also showed high sensitivity, with severe responses at response thresholds lower than many other species but possibly less sensitive than beaked whales. Humpback whales generally had fewer scored responses overall, and responses were scored with lower severity; these responses were typically also triggered at higher levels than for the minke and bottlenose whales. Compared to previously published data on pilot, killer, and sperm whales, humpback whales seemed to be a less-responsive species. They also seemed to habituate to the sonar when exposed repeatedly, and they reacted less strongly to exposure to naval sonar than to the playback of vocalizations from a natural predator.

Acknowledgments

We appreciate the effort of the entire 3S field research team during data collection. Thanks to Nina Nordlund (FFI) for support in the received level analysis on minke whale and Louise Burt (CREEM) for her statistical input. The project was funded by the Norwegian Ministry of Defense, the U.S. Office of Naval Research, the Netherlands Ministry of Defense, and DGA French Ministry of Defense. PLT acknowledges the support of the MASTS (The Marine Alliance for Science and Technology for Scotland) pooling initiative in the completion of this study. MASTS is funded by the Scottish Funding Council (Grant Reference HR09011) and contributing institutions. Both CH and LB were financially supported by the U.S. Office of Naval Research Grant N00014-12-1-0204, under the project entitled Multi-study Ocean acoustics Human effects Analysis (MOCHA).

Literature Cited

- Ainslie, M. A. (2010). *Principles of sonar performance modeling*. Chichester, UK: Springer-Praxis. <http://dx.doi.org/10.1007/978-3-540-87662-5>
- Alves, A., Antunes, R., Bird, A., Tyack, P., Miller, P. J. O., Lam, F.-P. A., & Kvadsheim, P. H. (2014). Vocal matching of naval sonar signals by long-finned pilot whales (*Globicephala melas*). *Marine Mammal Science*, 30(3), 1248-1257. <http://dx.doi.org/10.1111/mms.12099>
- 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 (*Globicephala melas*). *Marine Pollution Bulletin*, 83(1), 165-180. <http://dx.doi.org/10.1016/j.marpolbul.2014.03.056>
- Au, W. W. L. (1993). *The sonar of dolphins*. New York: Springer Verlag. <http://dx.doi.org/10.1007/978-1-4612-4356-4>
- Balcomb III, K. C., & Claridge, D. E. (2001). A mass stranding of cetaceans caused by naval sonar in the Bahamas. *Bahamas Journal of Science*, 5, 2-12.
- Bartón, K. (2011). *MuMin: Multi-model inference. R package, Version 1*. Vienna, Austria: R Foundation for Statistical Computing. Available from <https://cran.r-project.org/web/packages/MuMin/index/html>.
- Cook, M. L. H., Varela, R. A., Goldstein, J. D., McCulloch, S. D., Bossart, G. D., Finneran, J. J., . . . Mann, D. A. (2006). Beaked whale auditory evoked potential hearing measurements. *Journal of Comparative Physiology A – Neuroethology Sensory Neural and Behavioural Physiology*, 192(5), 489-495. <http://dx.doi.org/10.1007/s00359-005-0086-1>
- Cox, T. M., Ragen, T. J., Read, A. J., Vos, E., Baird, R. W., Balcomb, K., . . . Benner, L. (2006). Understanding the impacts of anthropogenic sound on beaked whales. *Journal of Cetacean Research Management*, 7(3), 177-187.
- Cranford, T. W., & Krysl, P. (2015). Fin whale sound reception mechanisms: Skull vibration enables low-frequency hearing. *PLOS ONE*, 10(1), e0122298. <http://dx.doi.org/10.1371/journal.pone.0116222>
- Croll, D. A., Clark, C. W., Calambokidis, J., Ellison, W. T., & Tershy, B. R. (2001). Effect of anthropogenic low-frequency noise on the foraging ecology of *Balaenoptera* whales. *Animal Conservation*, 4(1), 13-27. <http://dx.doi.org/10.1017/S1367943001001020>
- Curé, C., Antunes, R., Alves, A. C., Visser, F., Kvadsheim, P. H., & Miller, P. J. O. (2013). Responses of male sperm whales (*Physeter macrocephalus*) to killer whale sounds: Implications for anti-predator strategies. *Scientific Reports*, 3(1579). <http://dx.doi.org/10.1038/srep01579>
- Curé, C., Antunes, R., Samarra, F., Alves, A. C., Visser, F., Kvadsheim, P. H., & Miller, P. J. O. (2012). Pilot whales attracted to killer whale sounds: Acoustically-mediated interspecific interactions in cetaceans. *PLOS ONE*, 7(12), e52201. <http://dx.doi.org/10.1371/journal.pone.0052201>
- Curé, C., Sivle, L. D., Visser, F., Wensveen, P., Isojunno, S., Harris, C., . . . Miller, P. J. O. (2015). Predator sound playbacks reveal flight responses in a flight strategist baleen whale. *Marine Ecology Progress Series*, 526, 267-282. <http://dx.doi.org/10.3354/meps11231>
- 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(4), 452-472. <http://dx.doi.org/10.1578/AM.35.4.2009.452>
- Deecke, V. B., Slater, P. J. B., & Ford, J. K. B. (2002). Selective habituation shapes acoustic predator recognition in harbour seals. *Nature*, 420(6912), 171-173. <http://dx.doi.org/10.1038/nature01030>
- DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone, E. A., . . . Tyack, P. L. (2013). First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. *Biology Letters*, 9(4), 20130223. <http://dx.doi.org/10.1098/rsbl.2013.0223>
- Dombestein, E., & Gjersøe, A. L. (2012). *LYBN 6.1. user manual* (FFI-rapport 2012/01463). Retrieved 25 September 2015 from <http://rapporter.ffi.no/rapporter/2012/01463.pdf>.
- Ellison, W. T., Southall, B. L., Clark, C. W., & Frankel, A. S. (2012). A new context-based approach to assess marine mammal behavioural responses to anthropogenic sounds. *Conservation Biology*, 26(1), 21-28. <http://dx.doi.org/10.1111/j.1523-1739.2011.01803.x>
- Finneran, J. J., Houser, D. S., Mase-Guthrie, B., Ewing, R. Y., & Lingenfelter, R. G. (2009). Auditory evoked potentials in a stranded Gervais' beaked whale (*Mesoplodon europaeus*). *The Journal of the Acoustical Society of America*, 126(1), 484-490. <http://dx.doi.org/10.1121/1.3133241>
- Frantzis, A. (1998). Does acoustic testing strand whales? *Nature*, 392(6671), 29.
- Frid, A., & Dill, L. (2002). Human-caused disturbance stimuli as a form of predation risk. *Conservation Ecology*, 6(1), 11.
- Fristrup, K. M., Hatch, L. T., & Clark, C. W. (2003). Variation in humpback whale (*Megaptera novaeangliae*) song length in relation to low-frequency sound broadcasts. *The Journal of the Acoustical Society of America*, 113(6), 3411-3424. <http://dx.doi.org/10.1121/1.1573637>
- Gannon, D. P., Barros, N. B., Nowacek, D. P., Read, A. J., Waples, D. M., & Wells, R. S. (2005). Prey detection by bottlenose dolphins, *Tursiops truncatus*: An experimental test of the passive listening hypothesis. *Animal Behaviour*, 69, 709-720. <http://dx.doi.org/10.1016/j.anbehav.2004.06.020>
- Goldbogen, J. A., Southall, B. L., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Hazen, E. L., . . . Tyack, P. L. (2013). Blue whales respond to simulated mid-frequency military sonar. *Proceedings of the Royal Society B: Biological Sciences*, 280(1765). <http://dx.doi.org/10.1098/rspb.2013.0657>

- Harris, C. M., Sadykova, D., DeRuiter, S. L., Tyack, P. L., Miller, P. J. O., Kvadsheim, P. H., . . . Thomas, L. (2015). Dose response severity functions for acoustic disturbance in cetaceans using recurrent event survival analysis. *Ecosphere* (In review).
- Hooker, S. K., & Baird, R. W. (1999). Deep-diving behaviour of the northern bottlenose whale, *Hyperoodon ampullatus* (Cetacea: Ziphiidae). *Proceedings of the Royal Society of London B*, 266, 671-676. <http://dx.doi.org/10.1098/rspb.1999.0688>
- Houser, D. S., Helweg, D. A., & Moore, P. W. B. (2001). A bandpass filter-bank model of auditory sensitivity in the humpback whale. *Aquatic Mammals*, 27(2), 82-91.
- Isojunno, S., & Miller, P. J. O. (2015). Sperm whale response to tag boat presence: Biologically informed hidden state models quantify lost feeding opportunities. *ECOSPHERE*, 6(1), Article 6.
- Jefferson, T. A., Stacey, P. J., & Baird, R. W. (1991). A review of killer whale interactions with other marine mammals: Predation to coexistence *Mammal Review*, 21(4), 151-180. <http://dx.doi.org/10.1111/j.1365-2907.1991.tb00291.x>
- Johnson, M. P., & Tyack, P. L. (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering*, 28, 3-12. <http://dx.doi.org/10.1109/JOE.2002.808212>
- Kleinbaum, D. G., & Klein, M. (2005). *Survival analysis: A self-learning text* (3rd ed.). New York: Springer Science+Business Media.
- Kuningas, S., Kvadsheim, P. H., Lam, F-P. A., & Miller, P. J. O. (2013). Killer whale presence in relation to naval sonar activity and prey abundance in northern Norway. *ICES Journal of Marine Science*, 70(7), 1287-1293. <http://dx.doi.org/10.1093/icesjms/fst127>
- Kvadsheim, P. H., Lam, F-P. A., Miller, P. J. O., Alves, A. C., Antunes, R., Broccocelli, A., . . . Visser, F. (2009). *Cetaceans and naval sonar – The 3S-2009 cruise report* (FFI report 2009/01140). Retrieved 25 September 2015 from <http://rapporter.ffi.no/rapporter/2009/01140.pdf>.
- Kvadsheim, P. H., Lam, F-P. A., Miller, P. J. O., Doksæter, L., Visser, F., Kleivane, L., . . . Dekeling, R. (2011). *Behavioural response studies of cetaceans to naval sonar signals in Norwegian waters – 3S-2011 cruise report* (FFI report 2011/01289). Retrieved 25 September 2015 from <http://rapporter.ffi.no/rapporter/2011/01289.pdf>.
- Kvadsheim, P. H., Lam, F-P. A., Miller, P. J. O., Wensveen, P., Visser, F., Sivle, L. D., . . . Dekeling, R. (2012). *Behavioural responses of cetaceans to naval sonar signals in Norwegian waters – The 3S-2012 cruise report* (FFI-rapport 2012/02058). Retrieved 25 September 2015 from <http://rapporter.ffi.no/rapporter/2012/02058.pdf>.
- Kvadsheim, P. H., Lam, F-P. A., Miller, P. J. O., Wensveen, P., Visser, F., Sivle, L. D., . . . Dekeling, R. (2014). *Behavioural responses of cetaceans to naval sonar signals – The 3S-2013 cruise report* (FFI-rapport 2014/00752). Retrieved 25 September 2015 from <http://rapporter.ffi.no/rapporter/2014/00752.pdf>.
- Kvadsheim, P. H., Sivle, L. D., Miller, P. J. O., Wensveen, P., Tyack, P., Roos, M., . . . Dekeling, R. (2015). *The 3S2 experiments – Studying the behavioural effects of naval sonar on northern bottlenose whales, humpback whales and minke whales* (FFI-rapport 2015/01001). (In press). Retrieved 25 September 2015 from <http://rapporter.ffi.no/rapporter/2015/01001.pdf>.
- Littell, R., Milliken, G., Stroup, W., Wolfinger, R., & Schabenberger, O. (2006). *SAS for mixed models*. Cary, NC: SAS Institute Inc.
- Martin, S. W., Martine, C. R., Matsuyama, B. M., & Henderson, E. E. (2015). Minke whales (*Balaenoptera acutorostrata*) respond to navy training. *The Journal of the Acoustical Society of America*, 137(2533). <http://dx.doi.org/10.1121/1.4919319>
- Maybaum, H. L. (1993). Responses of humpback whales to sonar sounds. *The Journal of the Acoustical Society of America*, 94(1848). <http://dx.doi.org/10.1121/1.407710>
- McCordic, J. A., Todd, S. K., & Stevick, P. T. (2013). Differential rates of killer whale attacks on humpback whales in the North Atlantic as determined by scarification. *Journal of the Marine Biological Association of the United Kingdom*, 94(6), 1311-1315. <http://dx.doi.org/10.1017/S0025315413001008>
- McDonald, M. A., Hildebrand, J. A., Wiggins, S. M., & Ross, D. (2008). A 50 year comparison of ambient ocean noise near San Clemente Island: A bathymetrically complex coastal region off Southern California. *The Journal of the Acoustical Society of America*, 124(4), 1985-1992. <http://dx.doi.org/10.1121/1.2967889>
- Miller, P. J. O. (2006). Diversity in sound pressure levels and estimated active space of resident killer whale vocalizations. *Journal of Comparative Physiology A*, 192, 449-459. <http://dx.doi.org/10.1007/s00359-005-0085-2>
- Miller, P. J. O., Shapiro, A. D., & Deecke, V. B. (2010). The diving behaviour of mammal eating killer whales (*Orcinus orca* L., 1758): Variations with ecological not physiological factors. *Canadian Journal of Zoology*, 88, 1103-1112. <http://dx.doi.org/10.1139/Z10-080>
- Miller, P. J. O., Biassoni, N., Samuels, A., & Tyack, P. L. (2000). Whale songs lengthen in response to sonar. *Nature*, 405(6789), 903. <http://dx.doi.org/10.1038/35016148>
- Miller, P. J. O., Antunes, R., Alves, A. C., Wensveen, P., Kvadsheim, P. H., Kleivane, L., . . . Tyack, P. (2011). *The 3S experiments: Studying the behavioural effects of sonar on killer whales (Orcinus orca), sperm whales (Physeter macrocephalus), and long-finned pilot whales (Globicephala melas) in Norwegian waters* (Scottish Ocean Institute Technical Report SOI-2011-001). Retrieved 25 September 2015 from <http://soi.st-andrews.ac.uk/documents/424.pdf>.
- Miller, P. J. O., Antunes, R. N., Wensveen, P. J., Samarra, F. I. P., Alves, A. C., Tyack, P. L., . . . Thomas, L. (2014). Dose-response relationships for the onset of avoidance

- of sonar by free-ranging killer whales. *The Journal of the Acoustical Society of America*, 135(2), 975-993. <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., . . . Hooker, S. K. (2015). Northern bottlenose whales are highly sensitive to noise disturbance. *Proceedings of the Royal Society Open Science*. (In press).
- Miller, P. J. O., Kvadsheim, P. H., Lam, F-P. A., Wensveen, P. J., Antunes, R., Alves, A. C., . . . Sivle, L. D. (2012). The severity of behavioural 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. <http://dx.doi.org/10.1578/AM.38.4.2012.362>
- Moretti, D., Thomas, L., Marques, T., Harwood, J., Dilley, A., Neales, B., . . . Morrissey, R. (2014). A risk function for behavioural disruption of Blainville's beaked whales (*Mesoplodon densirostris*) from mid-frequency active sonar. *PLOS ONE*, 9(1). <http://dx.doi.org/10.1371/journal.pone.0085064>
- Morton, A. B., & Symonds, H. K. (2002). Displacement of *Orcinus orca* (L.) by high amplitude sound in British Columbia, Canada. *ICES Journal of Marine Science*, 59(1), 71-80. <http://dx.doi.org/10.1006/jmsc.2001.1136>
- Popper, A. N., & Ketten, D. (2008). Underwater hearing. In A. I. Bassbaum, K. Kaneko, G. M. Shepherd, & G. Westheimer (Eds.), *The senses – A comprehensive reference* (pp. 225-236). Oxford, UK: Elsevier Press. <http://dx.doi.org/10.1016/B978-012370880-9.00020-7>
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved 25 September 2015 from www.R-project.org.
- Richardson, W. J., Greene, C. R., Jr., Malme, C. I., & Thomson, D. H. (1995). *Marine mammals and noise*. San Diego: Academic Press.
- Simon, M., Johnson, M., & Madsen, P. T. (2012). Keeping momentum with a mouthful of water: Behaviour and kinematics of humpback whale lunge feeding. *Journal of Experimental Biology*, 215(21), 3786-3798. <http://dx.doi.org/10.1242/jeb.071092>
- Simpson, S. D., Meekan, M., Montgomery, J., McCauley, R., & Jeffs, A. (2005). Homeward sound. *Science*, 308(5719), 221. <http://dx.doi.org/10.1126/science.1107406>
- Sivle, L. D., Kvadsheim, P. H., Fahlman, A., Lam, F-P. A., Tyack, P. L., & Miller, P. J. O. (2012). Changes in dive behaviour during naval sonar exposure in killer whales, long-finned pilot whales, and sperm whales. *Frontiers in Physiology*, 3, Article 400. <http://dx.doi.org/10.3389/fphys.2012.00400>
- Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene, C. R., Jr., . . . Tyack, P. (2007). Marine mammal noise exposure criteria: Initial scientific recommendations. *Aquatic Mammals*, 33(4), 411-521. <http://dx.doi.org/10.1578/AM.33.4.2007.411>
- Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., & Henry, K. R. (1999). Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms. *The Journal of the Acoustical Society of America*, 106(2), 1134-1141. <http://dx.doi.org/10.1121/1.427121>
- Therneau, T. (2014). A package for survival analysis in S. *R package, Version 2.37-7*. Retrieved 25 September 2015 from <http://CRAN.R-project.org/package=survival>.
- Tubelli, A. A., Zosuls, A., Ketten, D. R., Yamato, M., & Mountain, D. C. (2012). A prediction of the minke whale (*Balaenoptera acutorostrata*) middle-ear transfer function. *The Journal of the Acoustical Society of America*, 132(5), 3263-3272. <http://dx.doi.org/10.1121/1.4756950>
- Tyack, P. L. (2000). Dolphins whistle a signature tune. *Science*, 289(5483), 1310-1311. <http://dx.doi.org/10.1126/science.289.5483.1310>
- Tyack, P. L. (2009). Human-generated sound and marine mammals. *Physics Today*, 62(11), 39-44. <http://dx.doi.org/10.1063/1.3265235>
- Tyack, P. L., Gordon, J., & Thompson, D. (2004). Controlled-exposure experiments to determine the effects of noise on marine mammals. *Marine Technology Society Journal*, 37, 39-51.
- Tyack, P. L., Zimmer, W. M. X., Moretti, D., Southall, B. L., Claridge, D. E., Durban, J. W., . . . Boyd, I. L. (2011). Beaked whales respond to simulated and actual navy sonar. *PLOS ONE*, 6(3). <http://dx.doi.org/10.1371/journal.pone.0017009>
- Visser, F., Miller, P. J. O., Antunes, R. N., Oudejans, M. G., Mackenzie, M. L., Aoki, K., . . . Tyack, P. L. (2014). The social context of individual foraging behaviour in long-finned pilot whales (*Globicephala melas*). *Behaviour*, 151(10), 1453-1477. <http://dx.doi.org/10.1163/1568539X-00003195>
- von Benda-Beckmann, A. M., Lam, F-P. A., Moretti, D. J., Fulkerson, K., Ainslie, M. A., van IJsselmuide, S., . . . Beerens, S. P. (2010). Detection of Blainville's beaked whales with towed arrays. *Applied Acoustics*, 71, 1027-1035. <http://dx.doi.org/10.1016/j.apacoust.2010.05.015>
- Williams, R., Erbe, C., Ashe, E., Beerman, A., & Smith, J. (2014). Severity of killer whale behavioural responses to ship noise: A dose-response study. *Marine Pollution Bulletin*, 79(1-2), 254-260. <http://dx.doi.org/10.1016/j.apacoust.2010.05.015>

Appendix A

Table A1. Severity score details for humpback whales (*Megaptera novaeangliae*) (mn). Exp id is the ID of the tag, and the exposure types are *no-sonar control* approaches, first (*sonar 1*) or second (*sonar 2*) sonar exposures, and killer whale sound and broadband noise playbacks (*kwPB* and *noise PB*, respectively). Each session may have several scores for different types of changes in behavior. For all the sonar exposures, the exact timing of the onset of the scored response was determined to assign a corresponding threshold for maximum received sound pressure level (SPL_{max}) and cumulative sound exposure level (SEL_{cum}). The duration of the response vs the duration of the exposure was determined to assign the correct severity score to the response in accordance with the criteria in Table 1. Occasionally, it was difficult to determine with certainty whether a change in behavior was a response to the experimental condition or a coincidental change; and in these cases, we cautiously scored it as a response but marked the score as a low confidence (*LC*) data point. Detailed data plots that were used as the basis for the scoring can be found in Kvadsheim et al. (2015).

Exp id	Score	Behavioral response and justification	Time of changepoint (UTC)	Duration of response/exposure (min)	SPL _{max} dB re 1 µPa	SEL _{cum} dB re 1 µPa's
mn11_157a no-sonar control	0	No response – No observed change in behavior.				
mn11_157a sonar 1	4	Brief avoidance – The animal turned away from the source briefly but returned to normal speed, directness, and course. The exact timing of the turn was determined from the pseudotrack.	09:23:27	5/10	102	97
mn11_157a sonar 1	0	No response – Unusual deep dive with wiggles but not considered a response because this type of dive was common in the baseline period.				
mn11_157a sonar 2	0	No response – Unusual deep dives with wiggles but not considered a response because this type of dive was common in the baseline period.				
mn11_157a sonar 2	0	No response – There was a small deviation in the horizontal track of the animal, but this was not considered to be a response.				
mn11_157a kwPB	5	Minor avoidance of sound source – The animal changed course by 90° away from the source but returned to its original course in the end of the playback		7.5/10		
mn11_157a kwPB	2	Brief change in dive profile – Wiggles in the dive in the middle of exposure.		5/10		
mn11_157a noise PB	6	Moderate avoidance of sound source – The animal turned 90° away from the source and returned to its original course.		20-30/10		
mn11_158ab no-sonar control	3	Minor change in locomotion – Change in direction and directness, and drop in speed, but no avoidance response.		6/10		
mn11_160ab no-sonar control	3	Minor change in locomotion – The animal slowed down during exposure.		4-5/10		
mn11_160ab sonar 1	0	No response – Change in dive profile, but similar changes also occurred frequently during baseline.				

mn11_160ab sonar 2	0	No response – No observed change in behavior.				
mn11_160ab kwPB	2	Moderate orientation response – There were no sightings of the animal during the exposure, but based upon the pseudotrack, we identified turns as being orientation responses.	15/15			
mn11_160ab kwPB	0	No response – Change in dive profile during exposure, but similar changes were also observed during the baseline period.				
mn11_160ab noise PB	5	Minor avoidance of sound source – The animal turned away from the source and increased speed.	10/15			
mn11_165def no-sonar control	3	Minor change in locomotion – The animal made a change in course which resulted in the animal moving parallel to the approaching ship, thus, not an avoidance. Based upon just one sighting.	9/10			
mn11_165def no-sonar control	6 LC	Moderate cessation of feeding – One animal stopped feeding; the other one continued. Low confidence.	12-14/10			
mn11_165def sonar 1	0	No response – The animal stopped feeding, but this also happened in two previous dives and in two dives later in the record.				
mn11_165def sonar 2	3	Minor change locomotion – The animal made a sharp turn towards the source. Two animals had tags.	7/10	21:03:38 (tag e) 21:03:34 (tag f)	106 (tag e) 106 (tag f)	100 (tag e) 100 (tag f)
mn11_165def kwPB	4	Moderate change in dive profile – The animal switched to very shallow diving.	20/15			
mn11_165def kwPB	6	Moderate avoidance of sound source – The animal turned away from the source for a time period approximately equal to the duration of the exposure.	15/15			
mn11_165def noise PB	0	No response – The animal moved away from the source already before the start of the playback.				
mn12_161ab no-sonar control	0	No response – A change in speed was observed, but similar changes were common in the baseline record. There was also a change in direction, but this occurred before the onset of the exposure.				
mn12_161ab sonar 1	6	Moderate avoidance – No visual sightings of the animal during exposure, but speed increased from pre- to post-exposure. The pseudotrack showed a change in direction away from the source.	30/10	01:00:28	132	132
mn12_161ab sonar 1	4	Moderate change in dive profile – Unusually long dive duration during exposure, followed by unusually shallow diving. Response started at ascent of the first dive during sonar.	13/10	00:59:59	122	118

mn12_161ab sonar 2 (no-ramp-up)	2	Brief change in dive profile – The animal aborted a dive at the start of the exposure, but the dives that followed appeared normal.	02:27:59	1/5	159	156
mn12_161ab noise PB	0	No response – The animal made a turn away from the source, but the pseudotrack revealed that this happened before the start of the playback.				
mn12_161ab kwPB	7 LC	Prolonged cessation of feeding – Low confidence because feeding in baseline was somewhat intermittent. Tag off 51 min after response started, but judged prolonged.		> 50/15		
mn12_161ab kwPB	6	Moderate avoidance – The animal made a sharp turn away from the source.		45/15		
mn12_161ab kwPB	3	Minor change in dive profile – Long abnormal dive at the start of the playback.		6/15		
mn12_164a no-sonar control	0	No response – No observed changes in behavior.				
mn12_164a sonar 1	6 LC	Prolonged change in inter-animal distribution – Group spacing of animals in the area changed. This response was consistent with other changes in movement, but exact timing was difficult to determine. It may have occurred before sonar onset. We consider this to be a response but have low confidence that it actually was in response to the sonar. We consider that the response started at the first surfacing after the start of the exposure.	00:01:49	100/10	142	138
mn12_164a sonar 1	6	Moderate avoidance – The animal changed direction away from the source but returned to previous course. The response was considered to start at the first surfacing after sonar onset.	01:01:49	20/10	142	138
mn12_164a sonar 1	6	Moderate cessation of feeding – Lunging (foraging) activity stopped soon after sonar onset. The response was considered to start during the first dive after the sonar onset when animal reached the average depth of lunges in previous dives.	00:01:13	20/10	125	119
mn12_164a sonar 1	5	Extended change in dive profile – Switched to shallow diving during exposure but returned to pre-exposure dive behavior. The response was considered to start at the beginning of the first shallow dive after the sonar onset.	01:01:27	20/10	125	119
mn12_164a sonar 2 (no-ramp-up)	0	No response – No observed changes in behavior.				
mn12_164a sonar 2 (no-ramp-up)	3	Minor change in locomotion – The animal turned towards the source and changed directness at sonar onset.	01:43:00	4/5	170	170
mn12_164a sonar 2 (no-ramp-up)	4	Brief cessation of feeding – Cessation of lunging coincided exactly with the duration of the sonar exposure. The onset of the response was during the first dive of the exposure when the animal reached the feeding depth of previous dives; the effect only lasted during that dive.	01:43:06	5/5	170	170
mn12_164b noise PB	0	No response – No observed changes in behavior.				

mn12_164a sonar 2 (no-ramp-up)	4	Brief cessation of feeding – Cessation of lunging coincided exactly with the duration of the sonar exposure. The onset of the response was during the first dive of the exposure when the animal reached the feeding depth of previous dives; the effect only lasted during that dive.	01:43:06	5/5	170	170
mn12_164b noise PB	0	No response – No observed changes in behavior.				
mn12_164b kwPB	7	Prolonged avoidance – The animal moved away from the source at increased speed. The speed increment lasted about 2.5 h.		150/15		
mn12_164b kwPB	5	Extended change dive profile – The onset was determined to be when the dive profile changed.		150/15		
mn12_170ab no-sonar control	0	No response – No observed changes in behavior.				
mn12_170ab sonar 1	7 LC	Prolonged cessation of feeding – Two animals were travelling together; both were tagged. One animal started feeding a few minutes before exposure started, but stopped feeding during exposure. The other animal did not feed at all in the period immediately before, during, or after exposure. This was scored to be a low confidence response because only one animal fed and feeding started immediately before exposure after a long period of not feeding. Such short duration feeding events were very rarely seen in the humpback whale records. This might indicate a low-quality food patch. The response was considered to start when the animal reached the depth of its previous lunge.	10:44:56 (tag a)	80/10	164 (tag a) 163 (tag a)	
mn12_170ab sonar 1	0	No response – There was a change in the dive profile of tag b (from wiggling to v-shaped dives), but this was common in baseline period.				
mn12_170ab sonar 2	0	No response – A small change in dive pattern was observed for tag a, but this started before sonar onset and, therefore, was not considered a response to sonar.				
mn12_170ab sonar 2	5	Minor avoidance – The two tagged animals changed heading and directness shortly during exposure. Exact change point occurred between visual observations and was determined based upon heading sensor on the tag.	12:34:17 (tag a) 12:34:17 (tag b)	7/10	142 (tag a) 140 (tag b)	145 (tag a) 141 (tag b)
mn12_170ab noise PB	0	No response – No observed changes in behavior.				
mn12_170ab noise PB	2	Multiple orientation responses – The two animals changed directness and heading several times, but these turns did not increase the distance to the source.		20/15		
mn12_170ab kwPB	4	Moderate change in dive profile – The two animals switched to shallower diving during playback.		10/15		
mn12_170ab kwPB	7	Prolonged avoidance – The two animals changed direction away from the source and increased speed.		~150/15		

mn12_170ab kwPB	7	Prolonged cessation of feeding – Both animals stopped lunge-feeding at the onset of the playback.		~60/15		
mn12_171ab no-sonar con- trol	0	No response – No observed changes in behavior.				
mn12_171ab sonar 1	5	Minor cessation of feeding – Short interruption of feeding (lunges) in the second half of the exposure. The response was considered to start at the beginning of the first dive after the last lunge dive.	18:44:12	2-3/10	179	179
mn12_171ab sonar 1	3	Minor change in dive profile – The dive profile changed from deep feeding dives to shallow dives in the end of the exposure, but the animal returned to normal dive behavior shortly after the exposure ended. The response started at the beginning of the first shallow dive.	18:44:09	5/10	179	179
mn12_171ab sonar 1	2	Multiple orientation responses – The animal changed direction several times, but these turns did not increase the distance to the source. This was determined from the pseudotrack and not from the visual track.	18:38:41	~6/10	137	134
mn12_171ab sonar 2	6	Moderate cessation of feeding – The animal conducted one feeding dive after sonar onset, stopped feeding, and then started again. The response was considered to start when the animal reached the depth of previous lunges.	19:53:58	15/10	146	142
mn12_171ab sonar 2	1	Brief orientation response – The animal turned briefly towards the source ship. Exact onset was determined from pseudotrack.	19:56:41	3/10	167	165
mn12_171ab kwPB	6	Moderate cessation of feeding – The animal stopped lunge feeding during exposure and started feeding again about 5 min after playback ended.		20/15		
mn12_171ab kwPB	6	Prolonged change in group distribution – Group size changed from two to a single animal during playback. The focal animal remained solitary for the rest of the tag record (> 6 h).		> 360/15		
mn12_171ab kwPB	6	Moderate avoidance of sound source – The animal turned away from the source but returned to its original course after about 20 min.		20/15		
mn12_171ab noise PB	0	No response – No observed changes in behavior.				
mn12_178a no-sonar con- trol	0	No response – A change in directness was observed, but this was also common prior to exposure and, therefore, was not considered a response.				
mn12_178a sonar 1	4 LC	Brief avoidance – The animal conducted a turn directed perpendicularly away from the source's line of approach during sonar exposure. However, the animal had an erratic horizontal movement before the exposure started, and, therefore, it was difficult to judge with confidence whether this turn was in response to sonar or just a random event. Therefore, it was given a low confidence response score.	07:27:59	2-3/10	143	141
mn12_178a sonar 2	0	No response – Some changes in horizontal movement were observed, but careful inspection of the pseudotrack confirmed that this was due to lunge feeding behaviour and was not likely a response to the sonar.				
mn12_179a no-sonar control 2	0	No response – No observed changes in behavior.				

mn12_179a sonar 1	2	Brief change in dive profile – The animal seemed to have extended the dive which started before sonar onset because this dive was unusually long relative to its maximum depth. The response was considered to start at sonar onset.	15:07:00	4/10	104	100
mn12_179a sonar 2	3	Minor change in dive profile – Very similar response as seen during the first sonar exposure. The animal seemed to have extended the dive which started before sonar onset. This unusual dive had a longer duration than during sonar 1 and, therefore, was scored to have a higher severity.	16:19:00	8-9/10	101	96
mn12_179a sonar 2	3	Minor change in locomotion – The animal made changes in direction and speed. The response started at the first change in direction based upon the pseudotrack.	16:20:54	5-10/10	143	142
mn12_180ab no-sonar control	4	Brief cessation of feeding – The animal stopped lunging for one dive.		3/10		
mn12_180ab sonar 1	6	Moderate cessation of feeding – The animal performed one normal dive with a feeding lunge after sonar onset, and then performed a very deep dive without lunges. The onset of response was considered to start at the beginning of the deep dive.	22:57:30	~60/10	165	160
mn12_180ab sonar 1	4	Moderate change in dive profile – Switched from typical shallow feeding dives to 20 m, and then performed a single very deep (> 100 m) dive. After this dive, the animal switched back to shallow diving but with much longer dive durations. The onset of the response was considered to start at the beginning of the deep dive.	22:57:30	~60/10	165	160
mn12_180ab sonar 1	5	Minor avoidance of sound source – The animal made a sharp turn and moved out of the path of the approaching sonar source.	22:55:00	7/10	94	88
mn12_180ab sonar 2	3	Minor change in dive profile – A similar response as seen during sonar 1. The animal performed a single unusual deep dive (100 m) during exposure before returning to normal shallow diving. The response started at the beginning of the deep dive.	00:57:44	20-30/10	146	143
mn12_180ab sonar 2	6	Moderate avoidance – The animal turned and started moving on a course that was perpendicular to the course of the sonar source. The time of the start of the sharp turn was determined from the pseudotrack.	00:55:24	20/10	127	122
mn12_180ab noise PB	0	No response – No observed changes in behavior.				
mn12_180ab kwPB	7	Prolonged avoidance – The animal changed direction 100° away from the playback source and continued on this new course until the end of the record.		>180/10		
mn12_180ab kwPB	3	Minor change in dive profile – The animal again conducted one unusually deep dive halfway into the exposure. This response was very similar to the responses seen during both sonar exposures.		9/10		

Table AII. Severity score details for the minke whale (*Balaenoptera acutorostrata*) (ba). Exp id is the ID of the tag, and the exposure types are no-sonar control and sonar. Each session may have several scores for different types of changes in behavior. For all the sonar exposures, the exact time of the onset of the scored response was determined to assign a corresponding maximum received sound pressure level (SPL_{max}) and cumulative sound exposure level (SEL_{cum}) thresholds of the response. The duration of the response vs the duration of the exposure was determined to assign the correct severity score to the response in accordance with the criteria in Table 1. Detailed data plots that were used as the basis for the scoring can be found in Kvadsheim et al. (2015).

Exp id	Score	Behavioral response and justification	Time of change point (UTC)	Duration of response / exposure (min)	SPL_{max} dB re 1 μ Pa	SEL_{cum} dB re 1 μ Pa's
ba11_180a no-sonar control	0	No response – The animal changed heading multiple times, but this was also common during the baseline period.				
ba11_180a no-sonar control	2	Brief change in dive profile – The animal conducted an unusual deep dive near the end of the exposure.		3/42		
ba11_180a sonar	2	Brief change in dive profile – The animal changed its dive pattern at sonar onset from very shallow diving (0 to 5 m) to diving into the sound channel at around 50 m depth. This behavior could be a possible attempt to improve the localization of the sound source.	02:20:00	5/70	83	83
ba11_180a sonar	7	Prolonged avoidance – The animal increased speed and swam directly away from the sound source throughout the rest of the exposure. Opportunistic visual observations of skim feeding at the surface before the start of the sonar exposure indicated that this response might also have involved a cessation of feeding. The start of the response was taken as the time of the first sighting at which the animal had increased speed and swam at a course directly away from the source.	02:29:15	> 180/70	146	155
ba11_180a sonar	8	Obvious aversion and sensitization – The animal was swimming at very high and increasing speed throughout the exposure period. About halfway into exposure, it changed its dive pattern to shallower dives and increased speed further, presumably to avoid the source more efficiently. This change in dive behavior was considered to be the onset of the response.	02:56:35	40/70	158	171

Table AIII. Severity score details for the northern bottlenose whale (*Hyperoodon ampullatus*) (ha). Exp id is the ID of the tag, and the exposure type is sonar. Each session may have several scores for different types of changes in behavior. For all the sonar exposures, we have also determined the exact time of the onset of the scored response in order to assign a corresponding maximum received sound pressure level (SPL_{max}) and cumulative sound exposure level (SEL_{cum}) thresholds of the response. The duration of the response vs the duration of the exposure was determined to assign the correct severity score to the response in accordance with the criteria in Table 1. Detailed data plots used as basis for the scoring can be found in Kvadsheim et al. (2015).

Exp id	Score	Behavioral response and justification	Time of change point (UTC)	Duration of response / exposure (min)	SPL _{max} dB re 1 µPa	SEL _{cum} dB re 1 µPa's
ha13_176 sonar	3	Minor change in locomotion – The animal swam towards the source and approached it at the very beginning of the exposure. Timing was determined from the pseudotrack.	09:57:58	8/35	86	88
ha13_176a sonar	5	Extended change in dive profile – The animal conducted an unusually deep dive during the exposure, deeper than any dive recorded in this species. This response was considered to start when the animal broke off the ascent of what appeared to be a shallow dive and began the decent to > 2,000 m depth.	10:06:14	> 420/35	122	124
ha13_176a sonar	7	Prolonged avoidance – The animal turned away from the source and continued on new track for > 7 h. Exact timing of the turn was determined from the pseudotrack.	10:06:25	> 420/35	130	128
ha13_176a sonar	8	Long-term avoidance of area – The animal had been relatively stationary in the same area before exposure but moved away and did not return to the exposed area within the record. In addition, a larger scale acoustic and visual survey of the larger exposed area showed that other animals in the area also evacuated for at least 6 h after the exposure.	10:06:25	> 420/35	130	128
ha13_176a sonar	7	Prolonged cessation of feeding – The animal was feeding (echolocating) prior to sonar onset, but never after. The start of this response was taken during the deep dive when the animal reached the average depth where echolocation started in previous feeding dives.	10:07:18	> 420/35	141	139
ha13_176a sonar	6	Extended cessation of vocal behavior – The animal was vocally active during entire tag record prior to sonar exposure, but turned completely silent for more than 7 h when the tag detached and the record ended. Onset of the response was the same as for cessation of feeding.	10:07:18	> 420/35	141	139

Appendix B: Supplementary Material for Statistical Analyses

Table B1. Backwards step-wise model selection for number of scored responses (*#Score*, top) and maximum score (*maxScore*, bottom). Type III Wald tests are shown for each candidate covariate with both test statistic (χ^2) and *p* value. Candidate covariates with the highest *p* value were dropped first (highlighted in **bold**). Retained models are shown on far right (*#Score*: Step 5; *maxScore*: Step 4). To diagnose any problems with the step-wise model selection in SAS, all combinations of explanatory variables for *#Score* were also compared using the Quasi-likelihood Information Criterion (QIC) (“dredge” function in package *MuMin* in R (Bartoń, 2011)). Both model selection methods arrived at the same final model (*#Score* ~ *kwPB* + *S*). *S* is all sonar exposures, *S2* is second sonar exposure, *no-ramp-up* is sonar exposures that did not include a ramp-up period, *PB* is all playbacks (both killer whale sound and noise playbacks), *PB2* is second playback, and *kwPB* is killer whale sound playbacks.

#Score	Step 1		Step 2		Step 3		Step 4		Step 5	
	108.9		109.3		107.6		108.9		108.3	
QIC	χ^2	<i>p</i> value	χ^2	<i>p</i> value	χ^2	<i>p</i> value	χ^2	<i>p</i> value	χ^2	<i>p</i> value
S	5.0	0.026	5.8	0.016	5.6	0.018	5.56	0.0184	5.37	0.0205
S2	2.7	0.100	2.7	0.100	2.7	0.100	2.94	0.0863		
PB	0.5	0.478								
PB2	0.9	0.344	0.2	0.622						
kwPB	36.8	0.000	27.6	0.000	31.5	0.000	31.46	0.0001	31.46	0.0001
no-ramp up	1.1	0.285	1.1	0.285	1.1	0.285				

maxScore	Step 1		Step 2		Step 3		Step 4	
	158.1		156.1		156.3		155.6	
QIC	Chis-q	<i>p</i> value	Chis-q	<i>p</i> value	Chis-q	<i>p</i> value	Chis-q	<i>p</i> value
S	3.6	0.059	3.6	0.059	6.7	0.010	6.8	0.009
S2	2.0	0.160	2.3	0.127	2.3	0.134		
PB	0.9	0.335	0.9	0.335				
PB2	5.2	0.022	5.2	0.022	5.5	0.019	5.8	0.016
kwPB	7.9	0.005	7.9	0.005	13.3	0.000	13.5	0.000
no-ramp up	0.0	0.984						

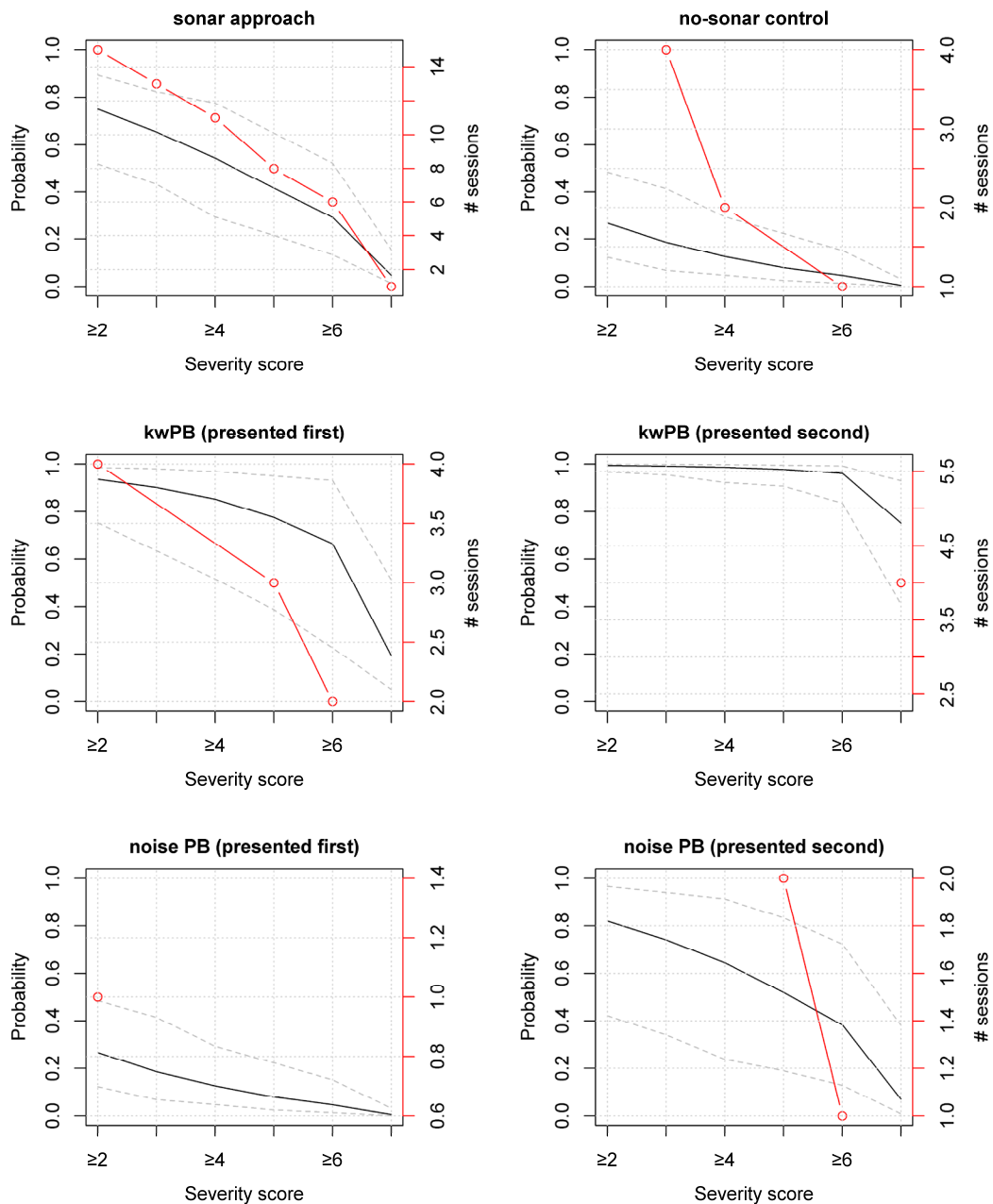


Figure B1. Fitted vs observed values from the model $\text{maxScore} \sim S + \text{kwPB} + \text{PB2}$ (Table B1). Solid black and dashed grey lines show mean and 95% confidence interval for the model predicted cumulative probability for a maximum score (i.e., probability of a score x or higher), while red connected dots show the cumulative number of sessions above the maximum score in the data. For example, the model predicted probability of a severity score 4 or higher during sonar approach was 0.54 (black line), while 11 out of the 20 sonar approaches (55%) were scored with a maximum severity of 4 or higher (red connected dots).