

## RESEARCH ARTICLE

# Lack of behavioural responses of humpback whales (*Megaptera novaeangliae*) indicate limited effectiveness of sonar mitigation

Paul J. Wensveen<sup>1,2,\*</sup>, Petter H. Kvadsheim<sup>3</sup>, Frans-Peter A. Lam<sup>4</sup>, Alexander M. von Benda-Beckmann<sup>4</sup>, Lise D. Sivle<sup>5</sup>, Fleur Visser<sup>6,7</sup>, Charlotte Curé<sup>8</sup>, Peter L. Tyack<sup>1</sup> and Patrick J. O. Miller<sup>1</sup>

## ABSTRACT

Exposure to underwater sound can cause permanent hearing loss and other physiological effects in marine animals. To reduce this risk, naval sonars are sometimes gradually increased in intensity at the start of transmission ('ramp-up'). Here, we conducted experiments in which tagged humpback whales were approached with a ship to test whether a sonar operation preceded by ramp-up reduced three risk indicators – maximum sound pressure level ( $SPL_{max}$ ), cumulative sound exposure level ( $SEL_{cum}$ ) and minimum source–whale range ( $R_{min}$ ) – compared with a sonar operation not preceded by ramp-up. Whales were subject to one no-sonar control session and either two successive ramp-up sessions (RampUp1, RampUp2) or a ramp-up session (RampUp1) and a full-power session (FullPower). Full-power sessions were conducted only twice; for other whales we used acoustic modelling that assumed transmission of the full-power sequence during their no-sonar control. Averaged over all whales, risk indicators in RampUp1 ( $n=11$ ) differed significantly from those in FullPower ( $n=12$ ) by  $-3.0$  dB ( $SPL_{max}$ ),  $-2.0$  dB ( $SEL_{cum}$ ) and  $+168$  m ( $R_{min}$ ), but not significantly from those in RampUp2 ( $n=9$ ). Only five whales in RampUp1, four whales in RampUp2 and none in FullPower or control sessions avoided the sound source. For RampUp1, we found statistically significant differences in risk indicators between whales that avoided the sonar and whales that did not:  $-4.7$  dB ( $SPL_{max}$ ),  $-3.4$  dB ( $SEL_{cum}$ ) and  $+291$  m ( $R_{min}$ ). In contrast, for RampUp2, these differences were smaller and not significant. This study suggests that sonar ramp-up has a positive but limited mitigative effect for humpback whales overall, but that ramp-up can reduce the risk of harm more effectively in situations when animals are more responsive and likely to avoid the sonar, e.g. owing to novelty of the stimulus, when they are in the path of an approaching sonar ship.

**KEY WORDS:** Behavioural effects, Hearing loss, Naval sonar, Baleen whale, Anthropogenic noise, Ramp-up

<sup>1</sup>Sea Mammal Research Unit, University of St Andrews, St Andrews, Fife KY16 9TS, UK. <sup>2</sup>Faculty of Life and Environmental Sciences, University of Iceland, Sturlugata 7, 101 Reykjavík, Iceland. <sup>3</sup>Maritime Systems, Norwegian Defence Research Establishment (FFI), NO-3191, Horten, Norway. <sup>4</sup>Acoustics and Sonar, The Netherlands Organisation for Applied Scientific Research (TNO), PO Box 96864, The Hague, 2509 JG, The Netherlands. <sup>5</sup>Marine Ecosystem Acoustics, Institute of Marine Research, PO Box 1870, Nordnes, 5817 Bergen, Norway. <sup>6</sup>Kelp Marine Research, Loniussstraat 9, 1624 CJ, Hoom, The Netherlands. <sup>7</sup>Behavioural Biology Group, Institute of Biology, Leiden University, PO Box 9505, 2300 RA, Leiden, The Netherlands. <sup>8</sup>Acoustics Group, CEREMA - DTerEst, F-67035, Strasbourg Cedex 2, France. \*Present address: Faculty of Life and Environmental Sciences, University of Iceland, Askja, Sturlugata 7, 101 Reykjavík, Iceland.

†Author for correspondence (pw234@st-andrews.ac.uk)

 P.J.W., 0000-0002-9894-2543

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Received 14 April 2017; Accepted 14 September 2017

## INTRODUCTION

Noise-induced hearing loss may affect individual survival, as acoustic communication in marine animals facilitates vital behaviours such as feeding and resting, and decisions about habitat selection via social information (Laiolo, 2010). Guidelines for human activities at sea that generate high-intensity sound currently recommend the use of operational mitigation measures designed to protect marine animals. One such mitigation measure is the gradual increase of source intensity prior to normal (full-power) operation, known as 'ramp-up' or 'soft-start'. Ramp-up is used by several navies during sonar exercises (Dolman et al., 2009) and is also common for other activities that involve high-intensity sources of sound, e.g. seismic surveys for oil and gas exploration (Nelms et al., 2016; Dunlop et al., 2016), and pile driving and detonation of explosions during offshore construction (Jefferson et al., 2009). The rationale behind ramp-up varies across sound producers, but the general argument is that for animals that would be close to the source on start-up, ramp-up prior to the start gives these animals time to move away before full-power transmission is reached. Ramp-up procedures for sonar operations are intended to mitigate against permanent hearing loss and other types of physiological effects in animals that are relatively close to the source, but might also protect against severe forms of behavioural disturbance (e.g. panic) in animals near a source that starts at full intensity.

Testing the effectiveness of ramp-up and other noise mitigation measures has been highlighted as a high priority for research in the scientific, conservation and regulatory communities (Southall et al., 2009; Shannon et al., 2016), indicating a clear societal need for such evaluations. Some operational mitigation procedures may prove expensive or affect the fidelity of naval combat training, making it unlikely that producers of underwater noise will adopt such procedures without specific regulations based upon empirical evidence of their conservation benefit. Already in 2009, a NATO working group on marine mammal risk mitigation identified testing the mitigation efficiency of operational methods such as ramp-up and visual-and-acoustic monitoring as a key data gap (R. Dekeling, personal communication). Thus far, however, the effectiveness of ramp-up for naval sonar has yet to be experimentally tested.

Using theoretical modelling, von Benda-Beckmann et al. (2014) found that ramp-up before normal sonar operation can be effective at reducing the number of simulated animals exposed to sound doses that are assumed to be high enough to cause temporary or permanent hearing loss. Important factors are the modelled relationship between acoustic dose and speed of the animals' avoidance response, as well as the ramp-up duration, sailing speed and time interval between the sonar pulses (von Benda-Beckmann et al., 2014). However, even though that study used relevant empirical input in their model, experimental confirmation of these predictions is still missing. It is therefore important to test the effectiveness of ramp-up in realistic conditions at sea.

**List of symbols and abbreviations**

$a$	gamma shape
$b$	gamma scale
$E$	single-pulse sound exposure
GEE	generalized estimating equation
PL	propagation loss
PTS	permanent hearing threshold shift
$r$	Pearson's correlation coefficient
$R_{\min}$	minimum source–whale range
SEL	single-pulse sound exposure level
SEL <sub>cum</sub>	cumulative sound exposure level of the session
SL	source level based on mean-square sound pressure
SL <sub>E</sub>	energy source level
SPL	sound pressure level
SPL <sub>max</sub>	maximum sound pressure level of the session
$T$	effective pulse duration
$t_0$	reference time=1 s
TTS	temporary hearing threshold shift
$Z$	test statistic for Wald test and Barnard's unconditional test
$\kappa$	concentration
$\mu$	mean
$\sigma$	standard deviation

In this study, the objective was to experimentally test the effectiveness of ramp-up of sonar transmitted by a moving ship on a cosmopolitan cetacean species, the humpback whale (*Megaptera novaeangliae* Borowski 1781). After being exploited during the whaling era, most humpback whale populations have been recovering relatively quickly compared with several other baleen whale sub-populations, many of which remain endangered or critically endangered (Thomas et al., 2016). The distribution of humpback whales during fitness-enhancing behaviours such as feeding and mating is concentrated at inshore and continental shelf waters (Christensen et al., 1992; Rosenbaum et al., 2014), which can overlap with that of naval sonar activity.

In the last two decades, several navies have started using lower-frequency ( $\leq 2$  kHz) active sonar systems for longer-range detections of submarines (Scott, 2015). These towed systems can ensonify larger areas of ocean as their signals are less directional and less influenced by absorption than higher-frequency signals, and thus have greater potential to affect humpback whales behaviourally or physiologically and to mask their acoustic communication.

Effectiveness of ramp-up of 1.3–2.0 kHz sonar was assessed in terms of whether risk of harm was less in whales presented with a full-power sequence preceded by a ramp-up sequence, i.e. the RampUp session, compared with whales presented with a full-power sequence preceded by a period without sonar, i.e. the FullPower session. In addition, the potential for short-term habituation or sensitization to the sonar exposure was investigated by presenting whales with two identical RampUp sessions. We analysed three risk indicators: maximum sound pressure level (SPL<sub>max</sub>), cumulative sound exposure level (SEL<sub>cum</sub>) and minimum source–whale range ( $R_{\min}$ ). In addition, effectiveness of ramp-up was assessed by identifying avoidance responses to the sound source and determining their influence on the risk indicators. We tested the specific scenario in which a whale was located directly in front of an approaching source ship and the sonar operation commenced relatively close to the whale, as that scenario likely poses the greatest risk of harm to whales in their natural environment.

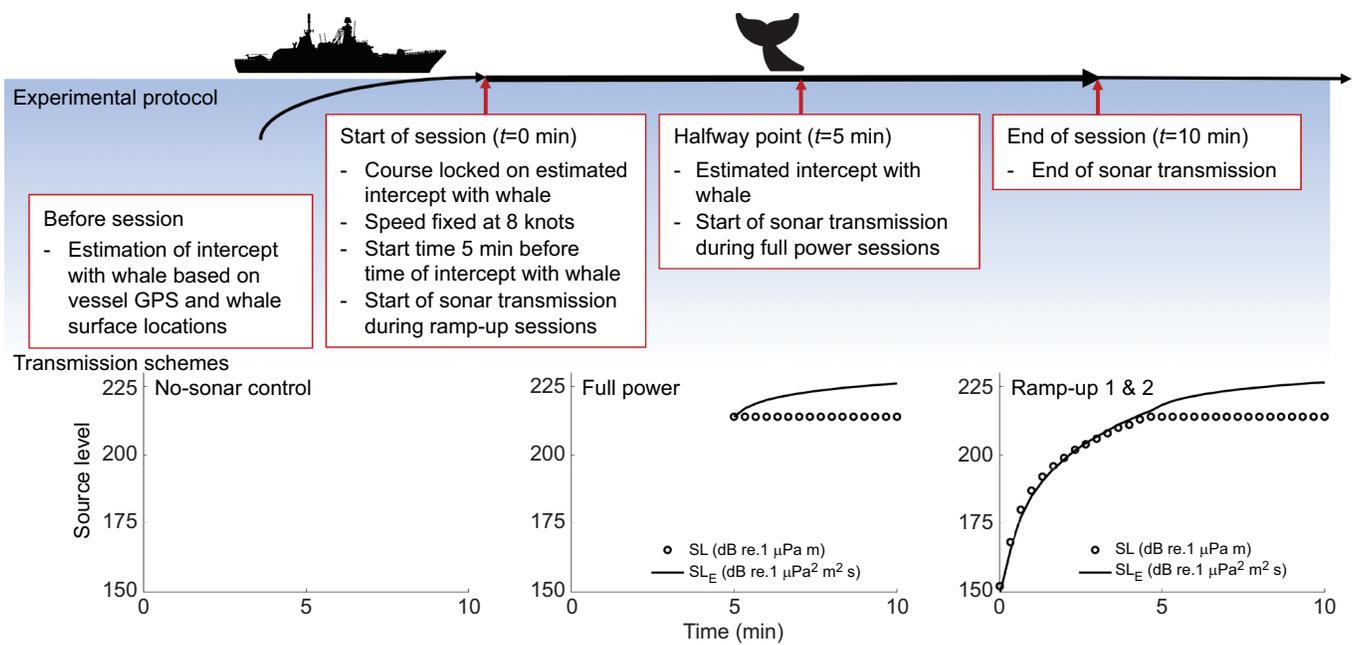
**MATERIALS AND METHODS****Data collection and experimental protocol**

Fieldwork was conducted in the Barents Sea between Bear Island and Spitsbergen in June 2011 and 2012 aboard the 55-m research vessel H. U. Sverdrup II (Kvadsheim et al., 2011). Details of the experimental protocols are described elsewhere (Kvadsheim et al., 2015) and summarized here. Humpback whales were detected visually from the flying bridge of the research vessel. After a whale was sighted, a tag boat was launched to deploy a multi-sensor tag [DTAG; Woods Hole Oceanographic Institution (WHOI), Woods Hole, MA, USA; Johnson and Tyack, 2003]. Tags were attached to the whale with suction cups using a cantilevered carbon fibre pole, or a pneumatic remote deployment system. DTAGs recorded sound (96 kHz; 16 bits), pressure, three-axial acceleration and three-axial magnetic field strength (50 Hz each). DTAGs also included a VHF beacon to aid visual tracking, and carried a small Fastloc-GPS logger (F2G 134A, Sirtrack, Havelock North, New Zealand) to record GPS surface locations. After a tagging period, groups of one or two tagged whales were tracked from an 8-m motorized boat at  $\sim 100$ –200 m. Observers recorded the distance and bearing to the focal whale, and number of whales in the group, for each new surfacing occurring at least 2 min after the last record. Pairs of whales were considered potential mother–calf pairs if they were composed of an adult and a smaller-sized individual that remained closely associated throughout the tracking record (Curé et al., 2015).

Experimental sessions started after 4 h (2012) or 8 h (2011) of baseline data were collected per whale. Each tagged whale was subject to three experimental sessions: first a no-sonar control, then two sonar sessions. Sessions had a 10-min duration each and were separated by 1 h or more. The no-sonar control was always conducted first to test how whales responded to the ship alone, before they heard sonar transmitted from the vessel. During no-sonar control the ship approached the tagged whale in the same way as during sonar sessions, but without transmitting sonar.

The source ship approached the whale at a speed of  $4.1 \text{ m s}^{-1}$  (8 knots) and on a predetermined straight intercept course during no-sonar control and sonar sessions (see Fig. 1 for a schematic of the navigational protocol). Both the line-of-approach and start time of the experimental session depended upon a prediction of the future track of the whale: each session started at 1.25 km from an estimated intercept with the tagged whale, as this was the distance covered during 5 min at  $4.1 \text{ m s}^{-1}$  sailing speed. Each session was conducted using two independent intercept calculator tools to support navigation during the vessel approach: MARIA GDK (Teleplan Globe, Lysaker, Norway), developed for the Norwegian Navy, and RU-tool (custom-built in MATLAB by P.J.W.). The surface positions of the whale, relayed via VHF radio by the observers on the tracking boat, and the GPS positions of the source ship were fed into the tools, which then calculated the time and position of the intercept and the best sailing path. The two tools used somewhat different logic: MARIA primarily used the automatic identification system signal from the tracking boat, which was usually around 100 m from the whale; and RU-tool predicted the future movements of the whale from the last sightings by applying different rules based on the whale's general movement pattern. The final course correction and start time of the session were determined only minutes before the start. Ultimate decisions on course changes and start of transmission were always made by the experiment coordinator based on all available information.

The two sonar sessions were either two identical ramp-up sessions ('RampUp1' and 'RampUp2') or a ramp-up session ('RampUp1') and a full-power session ('FullPower') without



**Fig. 1. The experimental protocol, including navigational protocol of the source ship and transmission schemes of the sonar.** Session names reflect both the stimulus presentation (no-sonar control, full-power and ramp-up) and order of presentation (ramp-up 1 and ramp-up 2). The navigational protocol was identical across experimental sessions but which transmission scheme was used depended on the type of session. The bottom panels show the source level in terms of RMS sound pressure (dotted line; dB re. 1  $\mu\text{Pa}$  m) as well as cumulative sound exposure (dB re. 1  $\mu\text{Pa}^2 \text{m}^2 \text{s}$ ).

ramp-up (Table 1). Ramp-up sessions consisted of a 5-min full-power operation preceded by a 5-min ramp-up, and full-power sessions consisted of a 5-min full-power operation preceded by 5 min without sonar transmissions. This design allowed us to test two distinct hypotheses: (1) that sonar operations preceded by ramp-up may be more effective at reducing risk of harm compared with sonar operations that are not preceded by ramp-up, which was tested using the full-power session and first ramp-up session, and (2) that whales may habituate or become sensitised to the sonar signal, which was tested using the two successive ramp-up sessions. The reader is referred to ‘Discussion, Methodological considerations’ for a description of how this design was established.

The research vessel towed the sonar source (Socrates II, TNO, The Hague, The Netherlands) at 50 m depth using 250–300 m of

cable. The source transmitted a 1.3–2.0 kHz upsweep pulse at 20 s interval. The single-pulse source level (SL; ISO18405:2017) was gradually increased from 152 to 214 dB re. 1  $\mu\text{Pa}$  m in the first 5 min and then kept at 214 dB re. 1  $\mu\text{Pa}$  m during ramp-up sessions (Fig. 1). Full-power sessions had no sonar transmission in the first 5 min and SL=214 dB re. 1  $\mu\text{Pa}$  m in the last 5 min (Fig. 1). The pulse duration was 1 s, except for the first 5 min of the ramp-up session, in which it was 0.5 s. We shortened the pulse duration to 0.5 s with the premise that the resulting reduction in cumulative exposure should reduce risk of hearing loss (von Benda-Beckmann et al., 2014).

At present, different ramp-up schemes are in operation worldwide. We could only test one ramp-up experimentally; therefore, the ramp-up scheme was based upon considerations of

**Table 1. Details of the experiments with humpback whales (*Megaptera novaeangliae*), including the types of experimental session conducted and the group composition during experimental sessions**

Tag ID	Group ID	Date and time (UTC)	Tagging location ( $^{\circ}\text{N}$ , $^{\circ}\text{E}$ )	Experimental session <sup>‡</sup>			Group composition
				1st	2nd	3rd	
mn11_157a	1	5 June 2011 22:25:15	75.143, 14.632	No-sonar	Ramp-up	Ramp-up	Solitary
mn11_158a	2	7 June 2011 09:21:23	74.830, 16.613	No-sonar	–	–	Group of 4, no calf
mn11_160a	3	8 June 2011 22:54:00	74.611, 15.294	No-sonar	Ramp-up	Ramp-up	Mother–calf pair
mn11_165e	4	14 June 2011 13:40:20	78.084, 11.085	No-sonar	Ramp-up	Ramp-up	Mother–calf pair
mn11_165f*	4	14 June 2011 13:59:51	78.084, 11.811	No-sonar	Ramp-up	Ramp-up	Mother–calf pair
mn12_161a	5	9 June 2012 16:44:27	77.344, 11.191	No-sonar	Ramp-up	Full-power	Group of 1–3, no calf
mn12_164a	6	12 June 2012 17:13:13	77.477, 09.594	No-sonar	Ramp-up	Full-power	Solitary
mn12_170a	7	18 June 2012 03:31:42	77.627, 10.403	No-sonar	Ramp-up	Ramp-up	Pair, no calf
mn12_170b	7	18 June 2012 03:49:48	77.463, 11.680	No-sonar	Ramp-up	Ramp-up	Pair, no calf
mn12_171b	8	19 June 2012 12:21:37	79.030, 10.669	No-sonar	Ramp-up	Ramp-up	Solitary
mn12_178a	9	26 June 2012 00:28:20	74.862, 17.805	No-sonar	Ramp-up	Ramp-up	Group of 2–3, no calf
mn12_179a	10	27 June 2012 07:57:13	74.045, 20.681	No-sonar	Ramp-up	Ramp-up	Solitary
mn12_180b	11	28 June 2012 17:07:52	73.983, 20.409	No-sonar	Ramp-up	Ramp-up	Mother–calf pair

\*Excluded from statistical analyses; no horizontal positions recorded.

<sup>‡</sup>The second session was always a ramp-up session (‘RampUp1’) and the third session was either a ramp-up session (‘RampUp2’) or a full-power session (‘FullPower’).

operational relevance and a quantitative assessment of risk of hearing loss. In this quantitative assessment, we applied a model framework described elsewhere (von Benda-Beckmann et al., 2014) with parameter values adjusted for humpback whales (speed:  $1.0 \text{ m s}^{-1}$ ) and sonar source used (pulse length during ramp-up phase: 0.2 s; other parameters as described above; details in Kvadsheim et al., 2011). The final parameter values for ramp-up duration, SL increase and range, and pulse duration, shape and interval, used during the at-sea experiments, were assumed to maximize the probability of avoidance and minimize the risk of hearing loss.

The animal experiments reported here were carried out under permits issued by the Norwegian Animal Research Authority (permit no. S2011/38782), in compliance with the ethical use of animals in experimentation. The research protocol was approved by the University of St Andrews Animal Welfare and Ethics Committee and WHOI's Institutional Animal Care and Use Committee.

### Modelling of full-power outcomes

Full-power sessions were only conducted twice (whales mn12\_161a and mn12\_164a) and always as the third session (Table 1). Records in the FullPower dataset for 10 other whales were based on (1) modelling of sound exposures that would have been received by the whales if the source had been transmitting the full-power sequence during no-sonar control sessions and (2) observed source–whale proximities during these no-sonar control sessions. The behaviour of the whales was not altered when full-power outcomes were generated. Thus, the FullPower dataset consisted of two whales (two sessions) that were exposed to a full-power sequence and 10 whales (those with horizontal positions recorded; nine sessions) for which the no-sonar control was used. This approach of generating full-power outcomes was taken because the navigational protocol, with source–whale intercept at  $t=5$  min, in combination with the full-power transmission sequence, with first the sonar transmission at  $t=5$  min (Fig. 1), was determined to not give whales enough time to substantially affect the risk indicators by avoiding the sonar. We simulated a full-power session to statistically assess the effect that an avoidance response to the first sonar transmission could have on  $\text{SPL}_{\text{max}}$ , if it occurred (Appendix A). In this simulation, avoidance behavior was modelled by randomly drawing samples from empirical distributions (e.g. for avoidance speed, depth, accuracy of intercept) derived from all observations collected during experimental sessions. The simulation showed that there was a negligible chance that an avoidance response would substantially reduce  $\text{SPL}_{\text{max}}$  (Appendix A).

Use of full-power outcomes in this way was considered the most appropriate approach from an animal-welfare perspective because it (1) reduced the number of times the source suddenly was switched on at maximum level at relatively close range from the tagged animal and (2) allowed us to obtain information about habituation and sensitization by exposing animals to two identical subsequent ramp-up sessions instead of to one ramp-up session and one full-power session. The two experimental full-power sessions were conducted to provide information about potential effects of the sudden nearby onset of sonar on the magnitude of the behavioural response (see 'Discussion, Methodological considerations' for the rationale behind this approach).

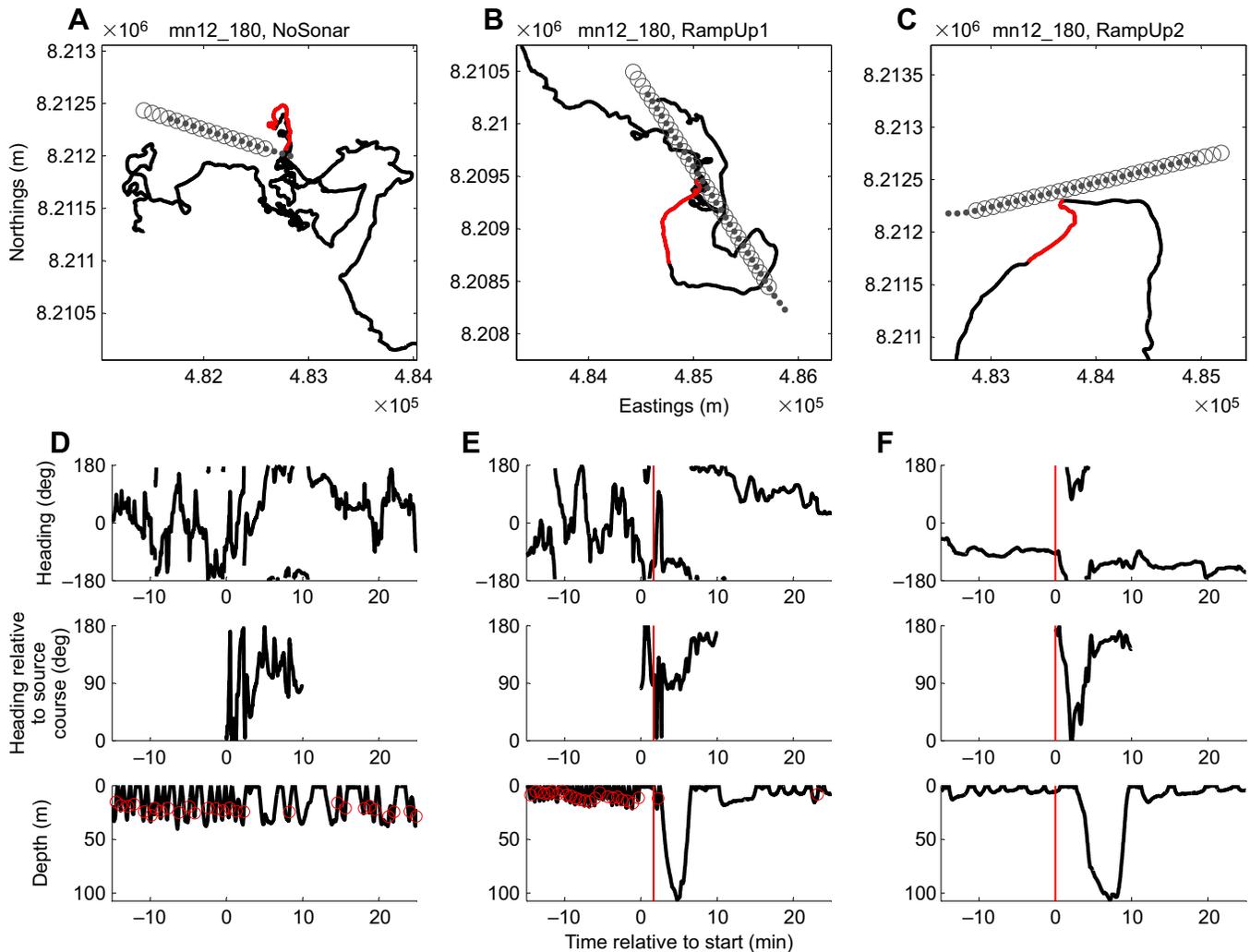
### Data analysis

We analyzed all experimental sessions to quantify three risk indicators – maximum sound pressure level ( $\text{SPL}_{\text{max}}$ ), cumulative sound exposure level ( $\text{SEL}_{\text{cum}}$ ) and minimum source–whale range ( $R_{\text{min}}$ ) –

which we could estimate from our at-sea treatments. Single-pulse sound pressure level (SPL) and sound exposure level (SEL) for sonar transmissions received on the DTAG were measured following the procedures in Miller et al. (2012). For definitions of SPL, SEL and SL we followed ISO 18405:2017 (<https://www.iso.org/standard/62406.html>).  $\text{SEL}_{\text{cum}}$  is the logarithm of the sum of the single-pulse sound exposures over all pulses in the session, and  $\text{SPL}_{\text{max}}$  is the maximum SPL of all pulses in the session. We combined SL with modelled acoustic propagation loss (PL) using ray-trace software (BELLHOP; Porter and Bucker, 1987) to obtain the received levels, i.e. the SPL and SEL of the sonar exposures received at the location of the whale, for all sonar sessions (Wensveen et al., 2015a). Sound speed profiles used for acoustic modelling were collected on-site using a conductivity-temperature-depth (CTD) profiler or an expendable bathythermograph, either during the experimental session or shortly after the session had ended (Kvadsheim et al., 2011). To correct for systematic discrepancy in the received levels (Fig. S1; Appendix B; see also Sivle et al., 2015; Wensveen et al., 2015a), we used the mean difference between the modelled received levels and the received levels measured on the DTAGs. The slant range between the three-dimensional position of the source and the three-dimensional position of the whale was calculated using high-resolution tracks of the whales, which were estimated from visual, GPS and tag data by Wensveen et al. (2015b). We applied a Monte Carlo simulation to propagate forward uncertainty in the whale's position into the modelled received levels. Further details on the received level analysis are provided in Appendix B.

For each session, a quantitative analysis to identify avoidance responses in the whale's horizontal track was conducted. This analysis aimed to replicate the definition of avoidance used by Sivle et al. (2015), who used expert identification of responses based on lower resolution tracks (i.e. only surface positions) for the same whales. Each track was downsampled to time steps of 20 s (the pulse interval) and then used to calculate whale heading. We identified all turns in the tracks and which of these could indicate avoidance. A turn was defined as a track segment in which the change in heading was either positive or negative and composed of one or several time steps. A turn was considered a potential avoidance response if it was  $\geq 90$  deg, occurred during a session and was followed by persistent movement away from the ship's trackline (defined as the whale moving at an average absolute heading between 0 and 100 deg relative to the course of the ship; Fig. 2, Fig. S2). To avoid misinterpretation of normal and coincidental changes in behaviour, the turn angle of each potential avoidance was compared with the empirical cumulative distribution of turn angles for all clockwise or counter-clockwise turns during 4 h of baseline data preceding the session. This baseline data set included turns during the period before the no-sonar control and, for sonar sessions, periods between experimental sessions. The duration of 4 h was chosen to reduce differences in the amount of data available per whale and variation in turn angles owing to whales switching behavioural state. Only turns in one direction were used to reduce autocorrelation in the baseline data set. Turns with a  $P$ -value of less than a Bonferroni-corrected significance level (0.05 divided by the number of tests on the experiment) were classified as avoidance.

We also assigned presence/absence of feeding just before or during the experimental session based upon the detection of lunges (Fig. 1D–F), i.e. feeding events in which the whale speeds up, engulfs a large volume of water and uses its baleen to filter prey (Goldbogen et al., 2008). 'Feeding' was assigned if at least one lunge was present in the time interval from 10 min before the start of the session until the onset of the response, or the end of the session if no response was scored. The feeding lunges used were identified by



**Fig. 2. Movements of a humpback whale mother–calf pair during a controlled exposure experiment.** Top panels show maps with the locations of tagged whale mn12\_180b (the mother) and locations of the sonar source during three experimental vessel approaches indicated: (A) no-sonar control session, (B) first ramp-up session and (C) second ramp-up session. The horizontal track of the whale (black line) is shown with the section corresponding to the experimental session highlighted (red line). The location of the sonar source (grey dots) and the location of the research vessel towing the source (grey circles) are shown for each individual sonar transmission. (D–F) Time-series data for the same experimental sessions, plotted as function of time relative to the start of the session: (1) whale heading relative to north, (2) whale heading relative to the course of the ship (where 0 deg represents whale movement in the same direction and 180 deg in the opposite direction) and (3) depth of the whale (black line) with indications when feeding events occurred (red circles). The onset of avoidance during the two ramp-up sessions is marked in the time-series plots (red vertical line). Similar plots were made for all experiments (Fig. S1).

Sivle et al. (2016) at depths  $>0.5$  m using a reliable detection algorithm (Simon et al., 2012).

### Statistical procedures

We compared the presence/absence of avoidance between no-sonar control sessions and ramp-up sessions (RampUp1 and RampUp2 combined) using a two-sided Barnard's unconditional test of superiority on a contingency table with avoidance (no response, response) and a null hypothesis on no effect of ramp-up session, relative to no-sonar control.

We were interested in examining whether the risk indicators ( $SPL_{max}$ ,  $SEL_{cum}$  and  $R_{min}$ ) changed across the three types of sonar session and across whales that avoided the source versus whales that did not. We used generalized estimating equations (GEEs), which are an extension of generalized linear models that accounts for correlated and clustered data by appropriately inflating the standard errors (Hardin and Hilbe, 2003). The dependent variable in the

models was the risk indicator.  $SPL_{max}$  and  $SEL_{cum}$  were modelled using Gaussian distributions and  $R_{min}$  using a gamma distribution.

All explanatory variables were factor covariates. We fitted models with session (FullPower, RampUp1, RampUp2) as the explanatory variable to examine the effects of transmission scheme and presentation order. As described above, we pooled 10 full-power outcomes that were based on no-sonar control sessions with two observed full-power sessions to create the FullPower dataset. This pooling assumed that the behaviour of the two whales during full-power sessions was unaffected by the previous two sessions. This assumption appears to be reasonable based on the movement tracks of the two exposed animals (Fig. S2E,F). All tagged whales were used in the analysis, resembling the real-world situation where animals encountering a sonar ship may or may not exhibit avoidance. We also fitted models with avoidance, ramp-up session (RampUp1, RampUp2) and their interaction as explanatory variables to examine the effects of avoidance and presentation order. Only ramp-up sessions

were used in this analysis because avoidance responses were not observed in full-power sessions.

Group ID was selected as blocking unit because data from both mn12\_170a and mn12\_170b were included and the behaviour of these associated whales may be not independent. All exposed whales were included in the statistical analysis except for whale mn11\_165f, as visual or GPS fixes were not recorded for this (non-focal) animal. The jackknife variance estimator was applied because the sandwich variance estimator can be biased for small sample sizes (Højsgaard et al., 2014). An independent correlation structure was used in all GEE analyses. We verified that three competing working correlation structures, i.e. exchangeable, AR1 and unstructured, did not improve the quasi-likelihood under the independence model (Hardin and Hilbe, 2003).

For inference purposes, we used prediction plots generated from the GEE models. The 95% confidence intervals (CIs) for the predictions were calculated using a parametric bootstrap on the GEE covariance matrix. The 2.5th and 97.5th percentiles were calculated from 5000 bootstrap iterations. We used two-sided Wald tests to assess differences between factor levels in models with session. For models with avoidance and ramp-up session, we used the output from the 5000 bootstraps to make comparisons between specific factor level combinations of interest, as the interaction term precluded the use of the standard Wald tests. We calculated the differences between predictions for different factor levels across all bootstraps as well as a 95% CI for these differences. We concluded that there was a significant increase or decrease in the risk indicator in cases where the upper and lower confidence bounds for the differences were exclusively positive or negative.

All statistical analyses were performed in R version 3.0.2 using R packages geepack (Højsgaard et al., 2014), MuMIn (Barton, 2015) and Barnard (Erguler, 2015). The raw data are available as Table S1.

## RESULTS

We successfully tagged 13 humpback whales, five in 2011 and eight in 2012 (Table 1). One whale was subject only to the no-sonar control because of a premature tag release. The remaining 12 whales were subject to three experimental sessions: first a no-sonar control, then two sonar sessions. These sonar sessions were either two consecutive ramp-up sessions (10 whales) or a ramp-up session followed by a full-power session (two whales; Table 1). The 13 whales were part of 11 independent groups of one or two tagged whales; on two occasions (mn11\_165ef and mn12\_170ab), two whales in the same group were tagged (Table 1). The number of groups tagged and exposed to sonar was based on a power analysis conducted after the first field season.

Whales were exposed to mean  $SPL_{max}=173$  dB re. 1  $\mu$ Pa (s.d. 4 dB) and  $SEL_{cum}=177$  dB re. 1  $\mu$ Pa<sup>2</sup> s (s.d. 3 dB) over all ramp-up and full-power sessions. The source passed at relatively close range to the whale: mean  $R_{min}$  was 299 m (s.d. 209 m) over all sessions. The experimental design, with the whale in front of an approaching

source, resulted in strong correlations among the three risk indicators:  $SPL_{max}$  and  $SEL_{cum}$  ( $r=+0.92$ ,  $P<0.001$ ),  $SPL_{max}$  and  $R_{min}$  ( $r=-0.93$ ,  $P<0.001$ ), and  $SEL_{cum}$  and  $R_{min}$  ( $r=-0.84$ ,  $P<0.001$ ).

Avoidance responses were identified in eight ramp-up sessions (nine whales) but never in the 11 no-sonar control sessions or two full-power sessions (Fig. S2). The presence of avoidance thus differed between stimulus presentations (Barnard's test:  $Z=2.8$ ,  $P=0.005$ ), suggesting that responses were caused by the sonar exposure and not the approaching ship. Between-whale and within-whale variations in the presence of avoidance were large; avoidance occurred during five of 10 RampUp1 sessions (five of 11 whales) and three of eight RampUp2 sessions (four of nine whales). Only one whale (mn12\_180b) avoided the source during both RampUp1 and RampUp2 (Fig. 1B,C). Of whales that avoided, four whales were in a feeding state and five whales were in a non-feeding state prior to the response. All avoidance responses were initiated when the source was approaching the whale (Fig. S2).

We investigated the effectiveness of ramp-up on humpback whales using statistical models with session (FullPower, RampUp1, RampUp2) as explanatory variable. FullPower consisted of observed sessions (two whales) and modelled outcomes (10 whales). The average differences in risk indicators between RampUp1 and FullPower were small but statistically significant (Fig. 2; Table 2). Differences for RampUp1 relative to FullPower were  $-3.0$  dB for  $SPL_{max}$  ( $Z=7.6$ ,  $P=0.006$ ),  $-2.0$  dB for  $SEL_{cum}$  ( $Z=11.6$ ,  $P<0.001$ ) and  $+168$  m for  $R_{min}$  ( $Z=8.8$ ,  $P=0.003$ ; Table 2). None of the contrasts with RampUp2 were statistically significant (Fig. 2, Table 2), and for all three risk indicators, the mean values in RampUp2 were closer to those in FullPower than to those in RampUp1.

Because effectiveness of ramp-up is mediated by the likelihood of inducing avoidance and by characteristics of the response (e.g. swim speed, heading), we also compared whales that avoided the sound source with whales that did not avoid. For RampUp1, we found statistically significant differences between these two categories:  $-4.7$  dB for  $SPL_{max}$ ,  $-3.4$  dB for  $SEL_{cum}$  and  $+291$  m for  $R_{min}$  (Fig. 2, Table 3). In contrast, for RampUp2, these differences were smaller overall and not statistically significant (Fig. 2, Table 3). Whales avoiding in RampUp2 received higher levels overall than whales avoiding in RampUp1 ( $+3.6$  dB for  $SPL_{max}$  and  $+4.1$  dB for  $SEL_{cum}$ ) and their minimum distance to the source averaged 209 m lower (Fig. 2, Table 3). These differences between whales that avoided were significant for SEL, but this was not the case for  $SPL_{max}$  and  $R_{min}$  (Table 3). Thus, while approximately the same percentage of whales avoided the sonar during RampUp1 compared with RampUp2, avoidance responses reduced received levels more effectively during RampUp1.

There were indications that a combination of behavioural state (feeding, non-feeding) and order of presentation influenced avoidance behaviour. The three whales in RampUp1 that initiated

**Table 2. Model statistics for the generalized estimating equation models with session (FullPower, RampUp1, RampUp2) as explanatory variable and maximum sound pressure level ( $SPL_{max}$ ), cumulative sound exposure level ( $SEL_{cum}$ ) or minimum source-whale range ( $R_{min}$ ) as the dependent variable**

Factor level 1	Factor level 2	$SPL_{max}$				SEL				$R_{min}$			
		Estimate (dB)	s.e.	Z	P	Estimate (dB)	s.e.	Z	P	Estimate (m)	s.e.	Z	P
RampUp1	FullPower	-3.0	1.1	7.6	0.006	-2.0	0.6	11.6	<0.001	168	57	8.8	0.003
RampUp2	FullPower	-0.9	1.5	0.4	0.542	0.2	1.0	0.03	0.866	41	72	0.3	0.567
RampUp2	RampUp1	2.1	1.9	1.2	0.284	2.2	1.2	3.3	0.071	-127	108	1.4	0.239

**Table 3. Results of the statistical comparisons between factor level combinations for the GEE models with avoidance (no response, response) and ramp-up session (RampUp1, RampUp2)\* as explanatory variables and maximum sound pressure level (SPL<sub>max</sub>), cumulative sound exposure level (SEL<sub>cum</sub>) or minimum source–whale range (R<sub>min</sub>) as the dependent variable**

Factor level combination 1	Factor level combination 2	SPL <sub>max</sub>			SEL			R <sub>min</sub>		
		Mean difference (dB)	95% CI		Mean difference (dB)	95% CI		Mean difference (m)	95% CI	
Response RampUp1	No response RampUp1	−4.7 <sup>‡</sup>	−8.9	−0.4	−3.4 <sup>‡</sup>	−6.0	−0.9	291 <sup>‡</sup>	32	555
Response RampUp2	Response RampUp1	3.6	−0.3	7.6	4.1 <sup>‡</sup>	1.2	7.1	−207	−503	85
Response RampUp2	No response RampUp2	−1.8	−5.1	1.6	0.1	−2.5	2.7	139	−49	328
No response RampUp2	No response RampUp1	0.7	−3.0	4.5	0.6	−1.8	3.1	−55	−202	84

\*This analysis only used ramp-up sessions because avoidance responses were not observed during full-power sessions.

<sup>‡</sup>Difference between factor level combinations was statistically significant ( $P < 0.05$ ).

avoidance from a non-feeding state received a mean SPL<sub>max</sub> = 167 dB re. 1 μPa (s.d. 5 dB), which was ~6 dB lower than other whales that avoided [feeding/RampUp1: 173 dB re. 1 μPa (s.d. 2 dB),  $n=2$ ; non-feeding/RampUp2: 174 dB re. 1 μPa (s.d. 2 dB),  $n=2$ ; feeding/RampUp2: 172 dB re. 1 μPa (s.d. 0 dB),  $n=2$ ]. Similar trends were observed for SEL<sub>cum</sub> and R<sub>min</sub>: whales that initiated avoidance from a non-feeding state in RampUp1 generally had a lower SEL<sub>cum</sub> and a greater R<sub>min</sub> than other whales that avoided.

## DISCUSSION

### Effectiveness of ramp-up for humpback whales

Ramp-up of 1.3–2.0 kHz sonar reduced risk, i.e. a decrease in SPL<sub>max</sub> and SEL<sub>cum</sub> and increase in minimum distance to the whale, of exposure to humpback whales, but only in the first sonar exposure session (RampUp1). The average reductions in risk for RampUp1 compared with FullPower were statistically significant, but relatively small (Fig. 2, Table 2). Animals exposed to an identical ramp-up sequence during two consecutive sessions did not exhibit similar risk reductions in the second session. However, unlike the differences in risk found between RampUp1 and FullPower, the differences in risk between RampUp1 and RampUp2 were not statistically significant, probably because of the smaller sample size and greater variation in responsiveness during RampUp2 (Table 2).

Changes in heading found to be avoidance responses were identified in approximately half of the whales (Fig. 3), which explains the small reductions in risk when averaged over all whales in the session. Whales that avoided the source during RampUp1 significantly reduced risk compared with whales that did not avoid during RampUp1, but also reduced their received level risk indicators compared with whales that avoided the source during RampUp2 (Fig. 2, Table 3). Thus, avoidance behaviour only reduced risk when whales had not been exposed to sonar 1 h before and the efficacy of ramp-up thus appeared to be influenced by the novelty of the stimulus. Such context-specificity is considered important for behavioural responses of humpback whales and other marine mammals (Ellison et al., 2012; Harris et al., in press). Furthermore, avoidance responses by the three whales that were in a non-feeding state during RampUp1 appeared to reduce risk more than avoidance responses by whales that were in a feeding state, suggesting that the behavioural responses of humpback whales to naval sonar varied with behavioural state, as in blue whales (*Balaenoptera musculus*; Goldbogen et al., 2013).

Five of the nine tagged whales exposed to two ramp-up sessions exhibited avoidance in only one session. This could be due to random variation in behavioural responses to man-made noise, common in baleen whales (McCauley et al., 2000; Goldbogen et al., 2013). An alternative explanation is that some of the whales were sensitized by the sonar whilst others habituated. The same humpback whales

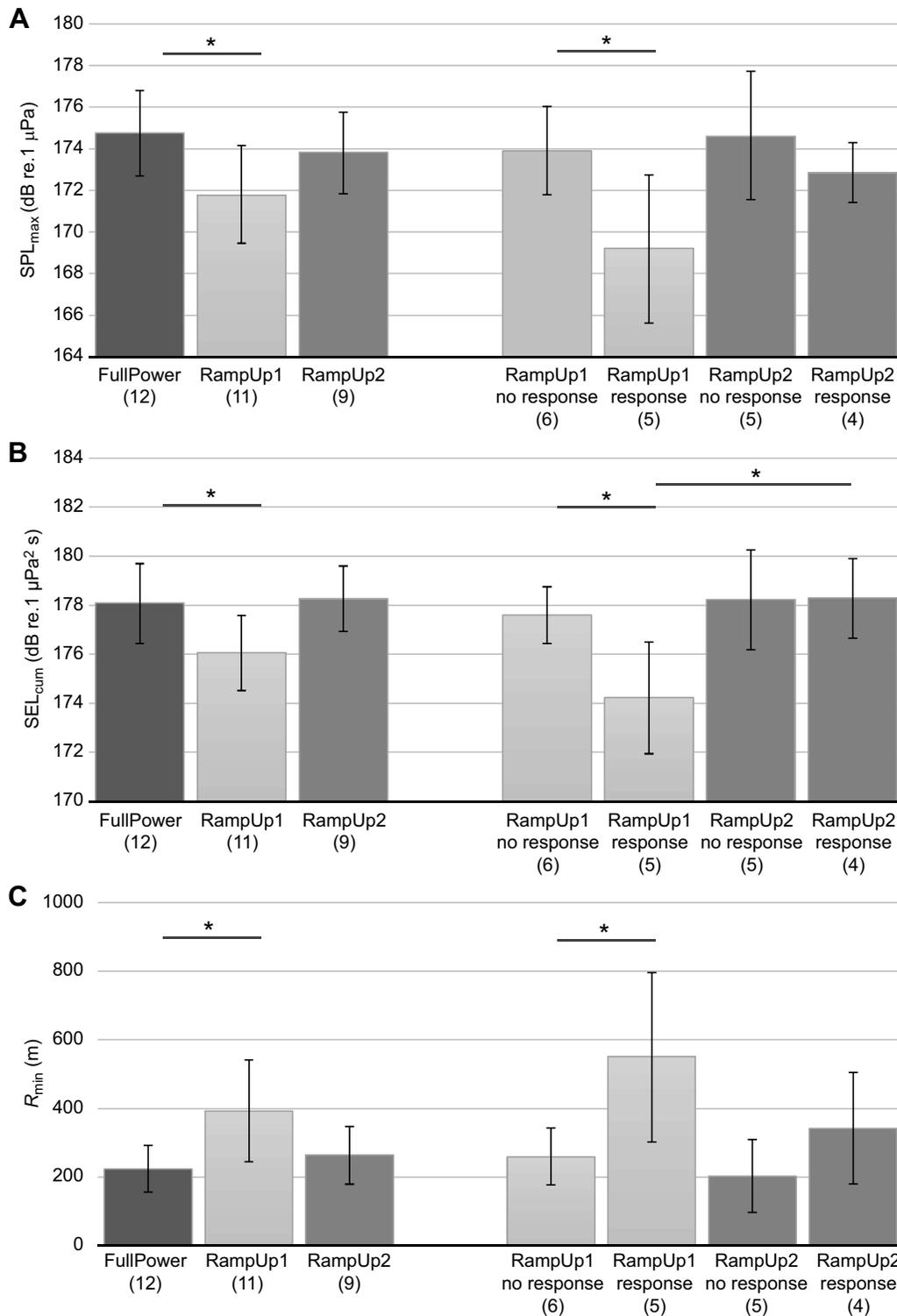
exhibited much stronger avoidance responses when killer whale sounds were played from a small drifting boat, at lower received levels than the sonar, documenting the whales' ability to respond more strongly, and indicating differences in level of perceived risk between the two stimuli (Curé et al., 2015; Sivle et al., 2015).

We did not analyze behavioural responses other than avoidance, but such responses (e.g. feeding cessations) have been reported in parallel studies of the same experiments (Sivle et al., 2015, 2016). These studies found a general tendency of habituation also in other aspects of the humpbacks' behavioural responses. A presentation-order effect might have occurred owing to the no-sonar control always being presented first. However, only one behavioural response (a brief cessation of feeding; Sivle et al., 2015) was observed during a no-sonar control session that was followed by an avoidance response in RampUp1, and this animal, mn12\_180b, did not habituate to the sonar during RampUp2. While a lack of response during the no-sonar control does not mean animals could not be sensitized by it, we consider this unlikely to have affected the overall pattern of habituation to the sonar.

The only tagged whale that avoided during both sonar sessions was a mother (mn12\_180b) with a small calf. Whale mn12\_180b responded during both ramp-up sessions with an unusual three-dimensional avoidance response, which included a descent to >100 m depth (Fig. 2E,F). The calf of mn12\_180b was substantially smaller than the other two calves (Fig. 4). Dive behaviour of the calf was not recorded as we did not tag it; however, this animal was always near its presumed mother when she surfaced. Young humpback whale calves produce very quiet calls while they are swimming (Videsen et al., 2017), so diving deep and increasing the horizontal distance to a masking sound source could help the mother to stay in contact with her calf. The observation is also consistent with the suggestion that mother–calf pairs are more likely to exhibit avoidance responses to man-made sounds that they are unaccustomed to (McCauley et al., 2000). We speculate that the mitigative effect of ramp-up may thus be higher for mothers with young calves compared with other groups. That the other two potential mother–calf pairs did not show the same strong response might be because these calves were older, which would be consistent with a reduction in parental investment in favour of foraging activity (Szabo and Duffus, 2008), or might also be caused by individual variation.

### Risk of hearing effects

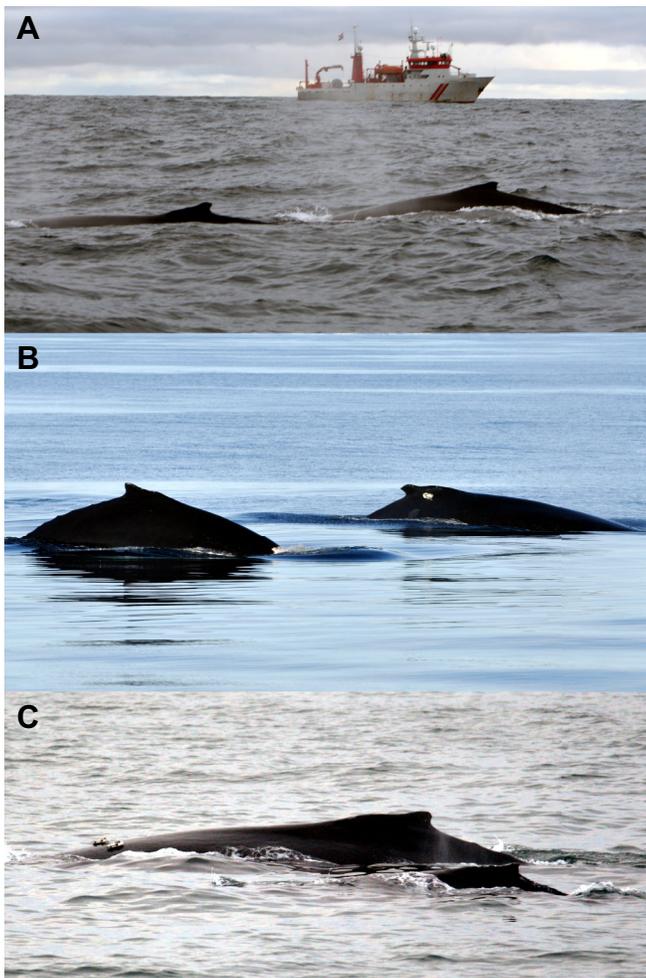
The source produced the relatively high received levels at the whale that were needed to invoke behavioural responses; the highest unweighted SEL<sub>cum</sub> was 183 dB re. 1 μPa<sup>2</sup> s (Table S1, Fig. S1). Direct measurements of temporary (TTS) or permanent hearing threshold shift (PTS) for humpback whales, or any other baleen



**Fig. 3. Generalized estimating equation (GEE) model predictions for the risk indicators measured.** Risk of harm was indicated using (A) maximum sound pressure level (SPL<sub>max</sub>), (B) cumulative sound exposure level (SEL<sub>cum</sub>) or (C) minimum source–whale range ( $R_{min}$ ). Significant ( $P < 0.05$ ) differences between factor levels or factor level combinations are indicated by an asterisk. Note that combination RampUp1/response was not statistically compared with combination RampUp2/no response (Table 3). Error bars indicate 95% confidence intervals obtained from a parametric bootstrap on the GEE covariance matrix using 5000 iterations. The number of whales ( $n$ ) is shown in parentheses below each bar.

whale, do not currently exist. The National Marine Fisheries Service has recently published updated risk thresholds, including an estimate of TTS onset (defined as a TTS of 6 dB) for low-frequency hearing specialist cetaceans, which includes humpback

whales, at frequency-weighted SEL<sub>cum</sub>=179 dB re. 1 μPa<sup>2</sup> s (NMFS, 2016). The NMFS (2016) auditory weighting function for these low-frequency cetaceans is 0 dB at 1.3–2.0 kHz; therefore, the unweighted and weighted SEL<sub>cum</sub> of the sonar are equal and the



**Fig. 4. Photographs of the potential humpback whale mother–calf pairs.** Groups of one or two tagged whales exposed to sonar included three potential mother–calf pairs: (A) mn11\_160a and an untagged adult (with the source ship in the background), (B) mn11\_165e and calf mn11\_165f, and (C) mn12\_180b and an untagged calf. Note that the calf of mn12\_180b was substantially smaller than the other two calves.

$SEL_{cum}$  calculated from the DTAG recordings (Fig. S1) can be directly compared with the risk thresholds of NMFS (2016). This comparison suggests that some of the subject animals may have experienced small amounts of TTS. However, the highest  $SEL_{cum}$  is still far (16 dB) below the PTS risk threshold of 199 dB re.  $1 \mu Pa^2 s$  (NMFS, 2016). Small threshold shifts in hearing sensitivity should have fully recovered after 1 h, and the 2.5–5% duty cycle of the sonar should have resulted in additional recovery of hearing during periods without direct path arrivals of sonar pulses, so it is unlikely that the reduced responsiveness during RampUp2 was the result of TTS from exposures during RampUp1. However, it is worthwhile to note that the reduction in  $SEL_{cum}$  achieved by the ramp-up protocol in this study for the responding animals is relevant, as it reduced the risk of TTS in some of those individuals based on these published risk criteria.

#### Methodological considerations

We used a sonar source that was previously tested as a prototype sonar system on operational Royal Netherlands Navy frigates. Its maximum SL falls in the low end of the range of operational tactical naval sonars (Ainslie, 2010). By design, ramp-up uses transmissions

at reduced SL to decrease the exposure at the animal, and therefore maximum SL was not a crucial factor in determining whether ramp-up is effective. However, theoretical modelling predicted that reductions in received level should be greater for sonar sources with higher maximum SLs, because animals are more likely to respond at greater distances (von Benda-Beckmann et al., 2016). The effectiveness of ramp-up is likely also determined by factors other than SL such as pulse interval, ship speed, whether the source is moving or stationary, and the animals' avoidance behaviour (von Benda-Beckmann et al., 2014, 2016; Wensveen et al., 2015a).

The initial plan was that each experiment would consist of one no-sonar control and two ramp-up sessions during the first season and one no-sonar control, one full-power and one ramp-up session during the second season. Full-power sessions would only be conducted after the first season because of ethical concerns that the sudden onset at maximum SL might cause severe behavioural disturbance. The order of the full-power and ramp-up session would be alternated per new experiment. Based on previous reports of responses of humpback whales to naval sonar (Maybaum, 1989) and other non-impulsive anthropogenic noise (Baker et al., 1983; McCauley et al., 1996), our expectation was that most avoidance responses would occur at SPLs of  $\sim 120$ – $140$  dB re.  $1 \mu Pa$ . However, exploratory analyses conducted after the first season indicated that the subject whales were relatively unresponsive. As the full-power protocol was determined to not give whales that avoided the source enough time to significantly affect the risk indicators, the decision was made to use modelled received levels that were based on hypothetical exposures using no-sonar control sessions. This adaptation of the original experimental design greatly increased the statistical power to detect effects. The lack of overt behavioural responses during ramp-up sessions in the first year also alleviated the initial welfare concerns about the full-power sessions. Therefore, we decided *a priori* to conduct two full-power sessions during the second season, to investigate potential effects of the sudden nearby onset of sonar on the magnitude of the behavioural response. While two observations would never have provided conclusive evidence, they could have provided relevant context to our results and indications of an effect of the sudden nearby onset.

#### Implications for conservation

Future marine conservation and management efforts critically rely upon our understanding of the acoustic environment as perceived by animals, the spatial and temporal patterns of disturbance, and immediate and chronic responses of animals to disturbance (Chen and Koprowski, 2015). Adaptive management of acoustic disturbance requires evaluation of mitigation methods such as ramp-up or mitigation zones, which require sonar shut-down (Dolman et al., 2009). The heterogeneity in behavioural responsiveness of baleen whales to sonar observed here and in other studies (e.g. Dunlop et al., 2013; Goldbogen et al., 2013) suggests that both between- and within-species variation should be considered when performing environmental risk assessments and evaluation of noise mitigation measures.

Our results represent the first experimental examination of ramp-up for naval sonar. The humpback whale was considered an appropriate model species for this purpose because the whales are relatively easy to find, tag and track. We found that gradually increasing the source intensity was not an effective method to reduce risk of physiological effects for humpback whales overall, because most whales did not exhibit very strong avoidance responses to the sonar signals. This is consistent with recent findings of migrating humpback whales exposed to airgun ramp-up in Australian waters (Dunlop et al., 2016).

However, more detailed analyses suggested that ramp-up of sonar reduces risk more effectively in situations in which animals might be more responsive, e.g. when animals have not been exposed recently, animals are in a non-feeding state, or a small calf is present. This suggests that ramp-up will have greater benefits for species that are more behaviourally responsive to sonar than humpback whales. These conclusions should not be extrapolated to ramp-up procedures for other anthropogenic sources, such as seismic airguns, as these sources and their ramp-up procedures are generally very different (Dunlop et al., 2016). Also, when animals have strong motivations not to move away from their current location, ramp-up may not be effective. Therefore, other operational mitigation procedures, e.g. warning sounds prior to sonar pulses (Nachtigall and Supin, 2013), should be considered in addition to ramp-up when mitigation strategies are designed.

The tagged whales were in the line-of-approach of the source ship during experiments and animals may have responded differently if the transmitting source was stationary or moving away. However, a whale directly in front of an approaching sonar ship is most likely to be at risk of physiological damage, and the purpose of ramp-up of sonar is specifically to reduce that risk for whales that happen to be directly in the path of the source ship before it starts a full-level sonar operation.

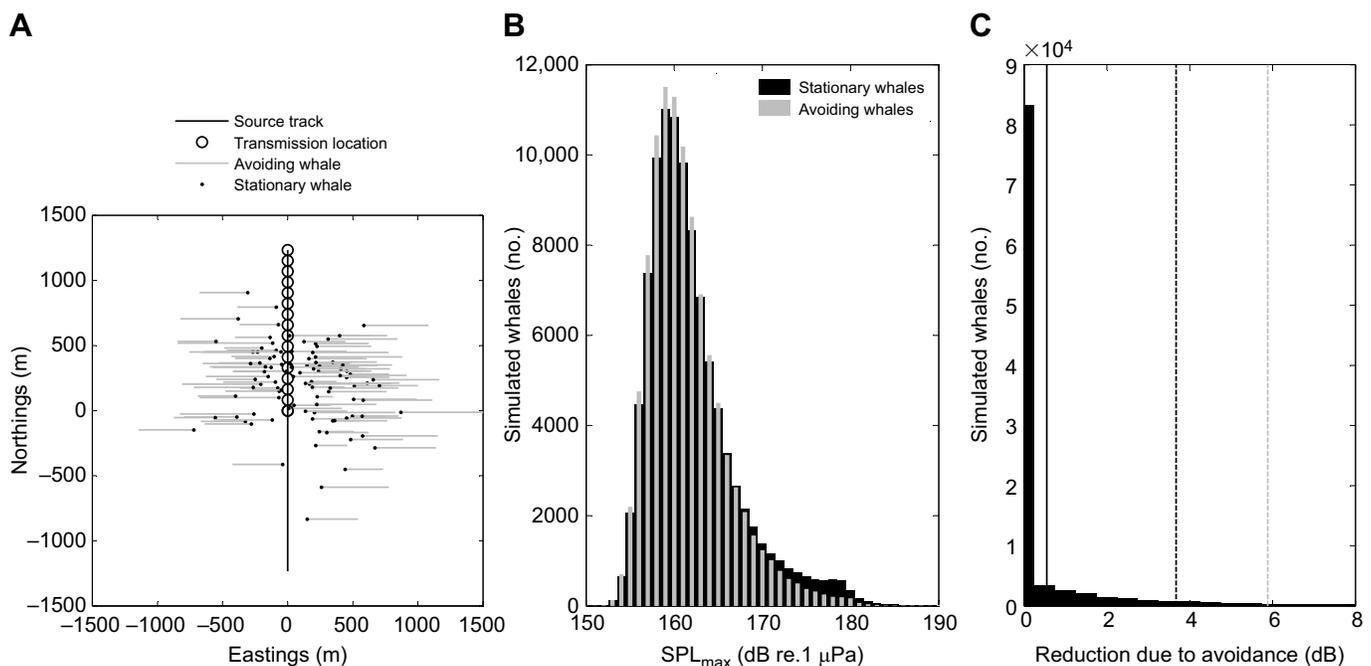
## APPENDIX A

### Simulation to assess the effect that avoidance could have on $SPL_{max}$ given the full-power protocol

A simulation was conducted to validate the assertion that whales could not reduce the risk indicators substantially by moving away in response to sonar during the full-power treatment. Specifically, we estimated the reduction in  $SPL_{max}$  that would result from an avoidance response to the first sonar transmission in the full-power session by comparing statistical distributions of  $SPL_{max}$  between

simulated whales that were stationary throughout the session and simulated whales that started moving in a direction perpendicular to the source track after the first transmission (Fig. A1). A fixed location was assigned to 50% of the simulated whales; the other 50% initiated avoidance and kept moving until sonar transmissions ceased. In each realization of the simulation, a simulated whale was assigned a different horizontal speed, depth and starting position. These values were randomly drawn from distributions fitted to actual whale positions observed at  $t=5$  min during no-sonar control and full-power sessions, and horizontal speed and depth of the whales during the observed avoidance responses. For whale starting positions, distance to the source was drawn from a truncated normal distribution with mean  $\mu=456$  m, s.d.  $\sigma=185$  m and lower bound=0 m, and the angle to the source was drawn from a von Mises distribution with  $\mu=20$  deg and concentration  $\kappa=0.89$ . Whale depth was drawn from a gamma distribution with shape  $a=0.5$  and scale  $b=37$  m, and horizontal avoidance speed was drawn from a truncated normal distribution with  $\mu=1.3$  m  $s^{-1}$ ,  $\sigma=0.5$  m  $s^{-1}$  and lower bound=0 m  $s^{-1}$ . The simulated source followed the navigational and transmission protocol of the full-power session (Fig. A1); horizontal speed=4.1 m  $s^{-1}$ , depth=50 m, and one pulse was transmitted at full power every 20 s between  $t=5$  and  $t=10$  min. We estimated  $SPL_{max}$  by assuming spherical spreading without absorption. The simulation only considered  $SPL_{max}$  because  $SEL_{cum}$  depends strongly upon  $SPL_{max}$  for close ship approaches. Simulations were run with 100,000 iterations.

The 75th, 95th and 97.5th percentiles of the reduction in  $SPL_{max}$  owing to avoidance were 0, 3.7 and 5.9 dB, respectively (Fig. A1). This indicated that there was a negligible chance during full-power sessions that a whale could substantially reduce its  $SPL_{max}$  by avoiding, given the experimental protocol of the full-power session and typical observed intercept distance, depth and avoidance speed of the whales.



**Fig. A1. Simulation to assess the effect that avoidance could have on  $SPL_{max}$  given the protocol of the full-power session.** (A) The track of the simulated source, with the source moving in northwards direction, and the locations for a subsample of simulated whales. (B) Monte Carlo distributions of  $SPL_{max}$  for all avoiding whales versus all stationary whales. (C) Monte Carlo distribution of the reduction in  $SPL_{max}$  caused by simulated avoidance responses, i.e. the difference between the two data sets shown in B. The mean and 95th and 97.5th percentiles are indicated by a dark solid line, dark dashed line and light dashed line, respectively.

One reason for this result was likely the large difference in speed between the source ( $4.1 \text{ m s}^{-1}$ ) and the receiver ( $\sim 1.3 \text{ m s}^{-1}$ ). All simulated whales avoided instantaneously and in a direction perpendicular to the source track (Fig. A1), which should maximize the reduction in received level (Wensveen et al., 2015a). However, actual observed avoidance responses had no instantaneous turn and generally less ‘optimal’ headings (often 0 deg relative to the course of the source; Fig. 2C, Fig. S2). Simulations were informed by parameters derived from avoidance responses during ramp-up sessions, and it is possible that the full-power protocol would, on average, induce stronger responses that would reduce risk of hearing loss more effectively. However, no avoidance responses were observed in the experimental two full-power sessions conducted, but both these whales were previously exposed.

## APPENDIX B

### Details of the received level analysis

We modelled acoustic PL using ray-trace software (BELLHOP; Porter and Bucker, 1987) to obtain the received levels for all sonar sessions, i.e. ramp-up sessions and simulated and observed full-power sessions. The acoustic model assumed a pressure-release sea surface and a bottom layer that was a flat, homogeneous fluid layer with constant acoustic properties. Bottom reflection coefficients were calculated using geo-acoustic parameters reported in Ainslie (2010) for the most prevalent sediment types at the exposure sites (Gurevich, 1995), in combination with the water sound speed and density, derived from CTD data, above the seafloor. For each experimental session, incoherent PL was modelled for a single two-dimensional slice with  $1 \times 1 \text{ m}$  grid resolution. Each slice was 4 km long, and its vertical dimension was taken as the mean sea floor depth, based on BODC (2010), between the source and whale at the start of the experimental session.

The modelled sound source was based on the properties of the real sonar source. We modelled PL at 1.6 kHz, i.e. the logarithmic middle of the band. The vertical directivity pattern of the real source at that frequency (3 dB beamwidth:  $\pm 40$  deg) was implemented. The range of beam take-off angles in the vertical plane was  $\pm 89$  deg. The number of traced Gaussian beams was 3200 (this number was automatically selected by BELLHOP). The modelled source was horizontally omnidirectional and was placed at the mean tow depth of the real source calculated over all pulses in the experimental session.

The energy source level ( $SL_E$ ) was derived from the SL based on mean-square sound pressure via  $SL_E = SL + 10 \log_{10}(T/t_0)$  dB, where  $t_0$  is 1 s. Because of a gradual onset and offset in the waveform, the effective duration  $T$  of a transmitted sonar pulse was 0.47 s and 0.93 s for the time intervals  $0 \leq t < 5$  min and  $5 \leq t \leq 10$  min, respectively. The received single-pulse SEL and single-pulse SPL were derived from the modelled PL as  $SEL = SL_E - PL$  and  $SPL = SL - PL$ , respectively.  $SEL_{cum}$  of the experimental session was then calculated by taking 10 times the 10-base log of the sum of the single-pulse sound exposures  $E$  received in the session, where  $E = 10^{SEL/10}$  dB  $\mu\text{Pa}^2 \text{ s}$ .  $SPL_{max}$  was taken as the maximum over all SPL in the session.

We applied a Monte Carlo method to propagate forward uncertainty in the three-dimensional position of the whale into the received levels ( $SPL_{max}$  and  $SEL_{cum}$ ). For whale depth we used a normal distribution centered at the depth measured by the DTAG at the time of the sonar transmission and s.d. = 1 m reflecting the precision of the pressure sensor (Johnson and Tyack, 2003). Wensveen et al. (2015b) used a Bayesian method to reconstruct probabilistic horizontal tracks for the humpback whales in this study

from the DTAG data (orientation, depth, flow noise), visual observations of range and bearing to the whale, and Fastloc-GPS positions. Therefore, we derived the uncertainty in horizontal source-whale range from the uncertainty in the spatial location of the whale track. Depending on whale and session, the s.d. for  $SPL_{max}$  ranged from 0 to 1.2 dB and the s.d. for  $SEL_{cum}$  ranged from 0 to 0.8 dB. Uncertainty in the horizontal track of the whale and, therefore, uncertainty in received  $SPL_{max}$  and received  $SEL_{cum}$  were larger for data sets without Fastloc-GPS positions (Wensveen et al., 2015b).

For all humpback whales exposed to sonar, Sivle et al. (2015) described how the received levels of the pulses were calculated from the acoustic recordings made by the DTAGs. Received levels were calculated using tag-specific mean sensitivity values based on calibration measurements conducted in the 1–2 months before each research trial in 2011, 2012 and 2013 (for more details, see Wensveen, 2016). On average, the modelled  $SPL_{max}$  was 5.8 dB lower than the  $SPL_{max}$  measured in the DTAG recordings, and this difference was 2.2 dB for  $SEL_{cum}$  (Fig. S1). This mean difference based on all sessions measured was added to the modelled received level, reflecting our higher confidence in the accuracy of the measurements. This correction only affected absolute values and did not affect the main results of this study, as they are based on differences in received level. Potential reasons for the discrepancy between the modelled and measured levels included the simplified assumptions of the acoustic modelling, limited environmental information and the difference in root mean square averaging time for the SPL calculations. In addition, the total  $SL_E$  calculated over all pulses in the ramp-up session was 0.43 dB higher than the total  $SL_E$  calculated over all pulses in the full-power session (Fig. 1). Therefore, we subtracted 0.43 dB from the  $SEL_{cum}$  for ramp-up sessions so that any potential statistical differences between RampUp and FullPower in this metric were caused by differences in behavior and not by differences in transmission protocols.

### Acknowledgements

We gratefully acknowledge the work of all students, scientists, engineers and ship’s crews of R/V H. U. Sverdrup II involved in phase 2 of the 3S Behavioural Response Study. We are especially grateful to L. Kleivane, S. Isojunno, C. Harris, S. DeRuiter, M. Roos, F. Samarra, M. Ainslie and the anonymous reviewers of different versions of the manuscript for their contributions to this work. Parts of this paper are reproduced from the PhD thesis of Paul Wensveen (University of St Andrews; Wensveen, 2016).

### Competing interests

The authors declare no competing or financial interests.

### Author contributions

Conceptualization: P.J.W., P.H.K., F.-P.A.L., P.L.T., P.J.O.M.; Methodology: P.J.W., P.H.K., F.-P.A.L., A.M.v.B.-B., P.L.T., P.J.O.M.; Software: P.J.W., A.M.v.B.-B.; Formal analysis: P.J.W., A.M.v.B.-B., L.D.S.; Investigation: P.J.W., P.H.K., F.-P.A.L., L.D.S., F.V., C.C., P.J.O.M.; Writing - original draft: P.J.W.; Writing - review & editing: P.J.W., P.H.K., F.-P.A.L., A.M.v.B.-B., L.D.S., F.V., C.C., P.L.T., P.J.O.M.; Supervision: P.J.O.M.; Project administration: P.J.O.M.; Funding acquisition: P.J.W., P.H.K., F.-P.A.L., P.L.T., P.J.O.M.

### Funding

This work was supported by the Office of Naval Research [N00014-10-1-0355], The Netherlands Ministry of Defence, Norwegian Ministry of Defence and French Ministry of Defence. Additional support was provided through a PhD studentship with matched funding from The Netherlands Ministry of Defence and the Natural Environment Research Council [NE/J500276/1 to P.J.W.]; and the MASTS (The Marine Alliance for Science and Technology for Scotland) pooling initiative to P.L.T. MASTS is funded by the Scottish Funding Council [HR09011] and contributing institutions.

### Supplementary information

Supplementary information available online at <http://jeb.biologists.org/lookup/doi/10.1242/jeb.161232.supplemental>.

## References

- Ainslie, M. A. (2010). *Principles of Sonar Performance Modeling*. Chichester, UK: Springer-Praxis.
- Baker, C. S., Herman, L. M., Bays, B. G. and Bauer, G. B. (1983). *The Impact of Vessel Traffic on the Behavior of Humpback Whales in Southeast Alaska: 1982 Season*. Seattle, WA: National Marine Fisheries Service, National Marine Mammal Laboratory.
- Barton, K. (2015). Package 'MuMIn: Multi-model inference' for R. Version 1.15.1.
- BODC (British Oceanographic Data Centre) (2010). The GEBCO\_08 Grid. Version 20100927. <http://www.gebco.net>.
- Chen, H. L. and Koprowski, J. L. (2015). Animal occurrence and space use change in the landscape of anthropogenic noise. *Biol. Conserv.* **192**, 315–322.
- Christensen, I., Haug, T. and Øien, N. (1992). Seasonal distribution, exploitation and present abundance of stocks of large baleen whales (Mysticeti) and sperm whales (*Physeter macrocephalus*) in Norwegian and adjacent waters. *ICES J. Mar. Sci.* **49**, 341–355.
- Curé, C., Doksæter Sivle, L., Visser, F., Wensveen, P. J., Isojunno, S., Harris, C. M., Kvadsheim, P. H., Lam, F.-P. A. and Miller, P. J. O. (2015). Predator sound playbacks reveal strong avoidance responses in a fight strategist baleen whale. *Mar. Ecol. Prog. Ser.* **526**, 267–282.
- Dolman, S. J., Weir, C. R. and Jasny, M. (2009). Comparative review of marine mammal guidance implemented during naval exercises. *Mar. Pollut. Bull.* **58**, 465–477.
- Dunlop, R. A., Noad, M. J., Cato, D. H., Kniest, E., Miller, P. J. O., Smith, J. N. and Stokes, M. D. (2013). Multivariate analysis of behavioural response experiments in humpback whales (*Megaptera novaeangliae*). *J. Exp. Biol.* **216**, 759–770.
- Dunlop, R. A., Noad, M. J., McCauley, R. D., Kniest, E., Slade, R., Paton, D. and Cato, D. H. (2016). Response of humpback whales (*Megaptera novaeangliae*) to ramp-up of a small experimental air gun array. *Mar. Pollut. Bull.* **103**, 72–83.
- Ellison, W. T., Southall, B. L., Clark, C. W. and Frankel, A. S. (2012). A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds. *Conserv. Biol.* **26**, 21–28.
- Erguler, K. (2015). Package 'Barnard' for R. Version 1.6. <https://cran.r-project.org/package=Barnard>
- Goldbogen, J. A., Calambokidis, J., Croll, D. A., Harvey, J. T., Newton, K. M., Oleson, E. M., Schorr, G. and Shadwick, R. E. (2008). Foraging behavior of humpback whales: kinematic and respiratory patterns suggest a high cost for a lunge. *J. Exp. Biol.* **211**, 3712–3719.
- Goldbogen, J. A., Southall, B. L., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Hazen, E. L., Falcone, E. A., Schorr, G. S., Douglas, A., Moretti, D. J. et al. (2013). Blue whales respond to simulated mid-frequency military sonar. *Proc. Biol. Sci.* **280**, 20130657.
- Gurevich, V. I. (1995). *Recent sedimentogenesis and environment on the Arctic Shelf of western Eurasia*. Norwegian Polar Institute Report 131. <https://hdl.handle.net/11250/173022>
- Hardin, J. W. and Hilbe, J. M. (2003). *Generalized Estimating Equations*. Boca Raton, FL, USA: Chapman and Hall/CRC Press.
- Harris, C. M., Thomas, L., Falcone, E. A., Hildebrand, J., Houser, D., Kvadsheim, P. H., Lam, F.-P. A., Miller, P. J. O., Moretti, D. J., Read, A. J. et al. (in press). Marine mammals and sonar: dose-response studies, the risk-disturbance hypothesis and the role of exposure context. *J. Appl. Ecol.*
- Højsgaard, S., Halekoh, U. and Yan, J. (2014). 'geepack': Generalized estimating equation package for R. Version 1.2-0. <https://cran.r-project.org/package=geepack>
- Jefferson, T. A., Hung, S. K. and Würsig, B. (2009). Protecting small cetaceans from coastal development: impact assessment and mitigation experience in Hong Kong. *Mar. Policy* **33**, 305–311.
- Johnson, M. P. and Tyack, P. L. (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE J. Ocean. Eng.* **28**, 3–12.
- Kvadsheim, P., Lam, F.-P., Miller, P., Doksæter, L., Visser, F., Kleivane, L., van Ijsselmuide, S., Samarra, F., Wensveen, P., Curé, C. et al. (2011). *Behavioural response studies of cetaceans to naval sonar signals in Norwegian waters - 3S-2011 cruise report*. FFI Report 2011/01289. Horten, Norway: Norwegian Defence Research Establishment (FFI). <http://www.ffi.no/no/Rapporter/11-01289.pdf>.
- Kvadsheim, P. H., Lam, F., Miller, P., Sivle, L. D., Wensveen, P., Roos, M., Tyack, P., Kleivane, L., Visser, F., van Ijsselmuide, S. et al. (2015). *The 3S2 experiments - studying the behavioral effects of naval sonar on northern bottlenose whales, humpback whales and minke whales*. Horten, Norway. FFI Report 2015/01001. Horten, Norway: Norwegian Defence Research Establishment (FFI). <http://www.ffi.no/no/Rapporter/15-01001.pdf>.
- Laiolo, P. (2010). The emerging significance of bioacoustics in animal species conservation. *Biol. Conserv.* **143**, 1635–1645.
- Maybaum, H. L. (1989). *Effects of a 3.3 kHz sonar system on humpback whales, Megaptera novaeangliae, in Hawaiian waters*. MSc thesis, University of Hawaii.
- McCauley, R. D., Cato, D. H. and Jeffery, A. F. (1996). *A Study of the Impacts of Vessel Noise on Humpback Whales in Hervey Bay*. The Queensland Department of Environment and Heritage, Maryborough Branch.
- McCauley, R. D., Fewtrell, J., Duncan, A. J., Jenner, C., Jenner, M.-N., Penrose, J. D., Prince, R. I., Adhitya, A., Murdock, J. and McCabe, K. (2000). Marine seismic surveys - a study of environmental implications. *Aust. Pet. Prod. Explor. Assoc. J.* **40**, 692–708.
- Miller, P. J. O., Kvadsheim, P. H., Lam, F.-P. A., Wensveen, P. J., Antunes, R., Alves, A. C., Visser, F., Kleivane, L., Tyack, P. L. and Doksæter Sivle, L. (2012). The severity of behavioral changes observed during experimental exposures of killer (*Orcinus orca*), long-finned pilot (*Globicephala melas*), and sperm (*Physeter macrocephalus*) whales to naval sonar. *Aquat. Mamm.* **38**, 362–401.
- Nachtigall, P. E. and Supin, A. Y. (2013). A false killer whale reduces its hearing sensitivity when a loud sound is preceded by a warning. *J. Exp. Biol.* **216**, 3062–3070.
- Nelms, S. E., Piniak, W. E. D., Weir, C. R. and Godley, B. J. (2016). Seismic surveys and marine turtles: An underestimated global threat? *Biol. Conserv.* **193**, 49–65.
- NMFS (National Marine Fisheries Service) (2016). *Technical Guidance for Assessing the Effects of Anthropogenic Sound on Marine Mammal Hearing: Underwater Acoustic Thresholds for Onset of Permanent and Temporary Threshold Shifts*. NOAA Technical Memorandum NMFS-OPR-55. Silver Spring, MD: US Department of Commerce, NOAA.
- Porter, M. B. and Buckner, H. P. (1987). Gaussian beam tracing for computing ocean acoustic fields. *J. Acoust. Soc. Am.* **82**, 1349–1359.
- Rosenbaum, H. C., Maxwell, S. M., Kershaw, F. and Mate, B. (2014). Long-range movement of humpback whales and their overlap with anthropogenic activity in the South Atlantic Ocean. *Conserv. Biol.* **28**, 604–615.
- Scott, R. (2015). Sound effects: low frequency active sonar comes of age. *IHC Jane's Navy Int.* **120**, Art. 1, 1–14.
- Shannon, G., McKenna, M. F., Angeloni, L. M., Crooks, K. R., Frstrup, K. M., Brown, E., Warner, K. A., Nelson, M. D., White, C., Briggs, J. et al. (2016). A synthesis of two decades of research documenting the effects of noise on wildlife. *Biol. Rev.* **91**, 982–1005.
- Simon, M., Johnson, M. and Madsen, P. T. (2012). Keeping momentum with a mouthful of water: behavior and kinematics of humpback whale lunge feeding. *J. Exp. Biol.* **215**, 3786–3798.
- Sivle, L. D., Kvadsheim, P. H., Curé, C., Isojunno, S., Wensveen, P. J., Lam, F.-P. A., Visser, F., Kleivane, L., Tyack, P. L., Harris, C. M. et al. (2015). Severity of expert-identified behavioural responses of humpback whale, minke whale, and northern bottlenose whale to naval sonar. *Aquat. Mamm.* **41**, 469–502.
- Sivle, L. D., Wensveen, P. J., Kvadsheim, P. H., Lam, F. P. A., Visser, F., Curé, C., Harris, C. M., Tyack, P. L. and Miller, P. J. O. (2016). Naval sonar disrupts foraging in humpback whales. *Mar. Ecol. Prog. Ser.* **562**, 211–220.
- Southall, B. L., Berkson, J., Bowen, D., Brake, R., Eckman, J., Field, J., Gisiner, R., Gregerson, S., Lang, W., Lewandoski, J. et al. (2009). *Addressing the Effects of Human-Generated Sound on Marine Life: An Integrated Research Plan for U.S. Federal Agencies*. Washington, DC: Joint Subcommittee on Ocean Science and Technology. <http://www.nmfs.noaa.gov/pr/pdfs/acoustics/jsost2009.pdf>.
- Szabo, A. and Duffus, D. (2008). Mother-offspring association in the humpback whale, *Megaptera novaeangliae*: following behaviour in an aquatic mammal. *Anim. Behav.* **75**, 1085–1092.
- Thomas, P. O., Reeves, R. R. and Brownell, R. L., Jr (2016). Status of the world's baleen whales. *Mar. Mamm. Sci.* **32**, 682–734.
- Videsen, S. K. A., Bejder, L., Johnson, M. and Madsen, P. T. (2017). High suckling rates and acoustic crypsis of humpback whale neonates maximise potential for mother-calf energy transfer. *Funct. Ecol.* **31**, 1561–1573.
- von Benda-Beckmann, A. M., Wensveen, P. J., Kvadsheim, P. H., Lam, F.-P. A., Miller, P. J. O., Tyack, P. L. and Ainslie, M. A. (2014). Modeling effectiveness of gradual increases in source level to mitigate effects of sonar on marine mammals. *Conserv. Biol.* **28**, 119–128.
- von Benda-Beckmann, A. M., Wensveen, P. J., Kvadsheim, P. H., Lam, F.-P. A., Miller, P. J. O., Tyack, P. L. and Ainslie, M. A. (2016). Assessing the effectiveness of ramp-up during sonar operations using exposure models. In *The Effects of Noise on Aquatic Life II* (ed. A. N. Popper and A. Hawkins), pp. 1197–1203. New York, NY: Springer.
- Wensveen, P. J. (2016). Detecting, assessing, and mitigating the effects of naval sonar on cetaceans. PhD thesis, University of St Andrews, St Andrews, UK.
- Wensveen, P. J., von Benda-Beckmann, A. M., Ainslie, M. A., Lam, F.-P. A., Kvadsheim, P. H., Tyack, P. L. and Miller, P. J. O. (2015a). How effectively do horizontal and vertical response strategies of long-finned pilot whales reduce sound exposure from naval sonar? *Mar. Environ. Res.* **106**, 68–81.
- Wensveen, P. J., Thomas, L. and Miller, P. J. O. (2015b). A path reconstruction method integrating dead-reckoning and position fixes applied to humpback whales. *Mov. Ecol.* **3**, 31.