Sperm whales reduce foraging effort during exposure to 1–2 kHz sonar and killer whale sounds

SAANA ISOJUNNO[1,5] CHARLOTTE CURÉ[2] PETTER HELGEVOLD KVADSHEIM[3] FRANS-PETER ALEXANDER LAM[4]
PETER LLOYD TYACK[1] PAUL JACOBUS WENSVEEN[1] AND PATRICK JAMES O'MALLEY MILLER[1]

[1]*Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife, KY168LB, UK*
[2]*Cerema, Dter Est, Acoustics Group, Strasbourg Cedex 2, F-67035, France*
[3]*Maritime Systems Division, Norwegian Defence Research Establishment (FFI), NO-3191, Horten, Norway*
[4]*Acoustics and Sonar, Netherlands Organisation for Applied Scientific Research (TNO), The Hague, 2509JG, The Netherlands*

Abstract. The time and energetic costs of behavioral responses to incidental and experimental sonar exposures, as well as control stimuli, were quantified using hidden state analysis of time series of acoustic and movement data recorded by tags (DTAG) attached to 12 sperm whales (*Physeter macrocephalus*) using suction cups. Behavioral state transition modeling showed that tagged whales switched to a non-foraging, non-resting state during both experimental transmissions of low-frequency active sonar from an approaching vessel (LFAS; 1–2 kHz, source level 214 dB re 1 μPa m, four tag records) and playbacks of potential predator (killer whale, *Orcinus orca*) sounds broadcast at naturally occurring sound levels as a positive control from a drifting boat (five tag records). Time spent in foraging states and the probability of prey capture attempts were reduced during these two types of exposures with little change in overall locomotion activity, suggesting an effect on energy intake with no immediate compensation. Whales switched to the active non-foraging state over received sound pressure levels of 131–165 dB re 1 μPa during LFAS exposure. In contrast, no changes in foraging behavior were detected in response to experimental negative controls (no-sonar ship approach or noise control playback) or to experimental medium-frequency active sonar exposures (MFAS; 6–7 kHz, source level 199 re 1 μPa m, received sound pressure level [SPL] = 73–158 dB re 1 μPa). Similarly, there was no reduction in foraging effort for three whales exposed to incidental, unidentified 4.7–5.1 kHz sonar signals received at lower levels (SPL = 89–133 dB re 1 μPa). These results demonstrate that similar to predation risk, exposure to sonar can affect functional behaviors, and indicate that increased perception of risk with higher source level or lower frequency may modulate how sperm whales respond to anthropogenic sound.

Key words: anthropogenic noise; behavioral budget; DTAG; functional state; naval sonar; northern Norway; Physeter macrocephalus; *risk–disturbance hypothesis; sperm whale; state-switching model; time series model*

When disturbed by humans, animals may reduce their investment of time and energy on fitness-enhancing activities, such as foraging or resting. If persistent, such behavioral effects might lead to increased population vulnerability (Gill et al. 2001, Frid and Dill 2002, NRC 2005, Beale 2007, Bejder et al. 2009, Berger-Tal et al. 2011, Ellison et al. 2012, Sih 2013). Costs to fitness-enhancing activities and subsequent life functions are critical to link changes in individuals' fitness to

impacts at a population level (NRC [National Research Council] 2005). Species with relatively slow life history traits and high parental investment, such as sperm whales (*Physeter macrocephalus*), have evolved strategies to maximize survival and can be expected to rely largely upon phenotypic plasticity, including learning, to cope with rapid environmental change and repeated exposure to anthropogenic stressors (Sih 2013).

Cetaceans currently face changes in their marine habitat such as introduction of anthropogenic noise, chemical and marine debris pollution, and exploitation of their prey species (Reeves 2003). Cetaceans rely upon sound for foraging, communication, and navigation and are therefore thought to be especially

vulnerable to anthropogenic noise (e.g., from vessel traffic, naval operations, seismic exploration; Southall et al. 2007, Tyack 2009). Effects of active naval sonar are of particular concern due to high source levels, potential wide spatial scale of impact, and stranding events of several species, especially beaked whale species, coincident with sonar exercises (Frantzis 1998, Balcomb and Claridge 2001, D'Amico et al. 2009, Filadelfo et al. 2009). Both behavioral and physiological mechanisms have been suggested as the causal link between sonar exposure and strandings (Cox et al. 2006), and behavioral changes have been reported at much lower levels of active sonar than would have been expected to cause injury (Tyack et al. 2011, Miller et al. 2012, 2014, DeRuiter et al. 2013*b*). Deep divers such as sperm whales and beaked whales also have natural high nitrogen saturation levels (de Quirós et al. 2012), and may respond to anthropogenic noise in a way that increases risk of decompression sickness (Kvadsheim et al. 2012, Fahlman et al. 2014).

Documented behavioral responses of cetacean species to naval sonar range from potentially benign changes in behavior, e.g., orientation responses of blue whales (*Balaenoptera musculus*) and sperm whales (Miller et al. 2012, Goldbogen et al. 2013) and vocal matching by false killer (*Pseudorca crassidens*) and pilot whales (*Globicephala* spp.) (DeRuiter et al. 2013*a*, Alves et al. 2014), to behavioral effects that may be more directly linked to fitness, such as avoidance and reduction of foraging in killer whales (*Orcinus orca*; Miller et al. 2012, 2014), beaked whales (Tyack et al. 2011, DeRuiter et al. 2013*b*), blue whales (Goldbogen et al. 2013), long-finned pilot whales (*Globicephala melas*; Antunes et al. 2014), and sperm whales (Miller et al. 2012). However, the biological consequences of such cetacean responses and the potential impact of cumulative exposures of naval sonar in combination with other anthropogenic noise sources remain poorly understood (Southall et al. 2007, Clark et al. 2009, Wright et al. 2011).

Behaviorally mediated costs of human disturbance are partly expected because animal species have evolved response and learning strategies to both familiar and novel stimuli that might indicate opportunity and/or risk (Sih 2013). Human presence or activity may be perceived as a form of predation risk, influencing individual's cost–benefit assessment of behavioral options ("risk–disturbance hypothesis"; Frid and Dill 2002). Because individuals weigh the cost–benefit of behavior change (e.g., leaving a high-quality patch) against perceived risk (Brown and Kotler 2004), the incurred costs can be used to indicate the magnitude of perceived risk in a given decision-making context (e.g., individual state, resource availability). Thus, playback of natural predator stimuli as a positive control can provide a useful yardstick for what constitutes a biologically significant response in behavioral response studies (BRS) of anthropogenic noise.

In this study, we quantified the relative time and energetic costs of behavioral responses of high-latitude male sperm whales (*Physeter macrocephalus*) to naval sonar exposures and natural predator stimuli. Sperm whale males are nearly twice the mass of females, and migrate from low-latitude breeding grounds to these highly productive waters (Best 1979, Teloni et al. 2008, Engelhaupt et al. 2009, Oliveira and Wahlberg 2013). In our study area, successful foraging is therefore likely to be particularly important to male fitness, including stores for migration, growth, and subsequent breeding success. Thus, "costs" could be defined in terms of foraging behavior: reduction in foraging time, reduction in foraging success, or increase in foraging costs. A state-based approach was applied to estimate costs within and across functional behavior states, such as foraging and resting (Isojunno and Miller 2015).

Tagged sperm whales were subject to controlled exposure experiments (CEE) including low-frequency active sonar (LFAS 1–2 kHz), mid-frequency active sonar (MFAS 6–7 kHz) transmissions from an approaching source, and playbacks of mammal-eating killer whale (*Orcinus orca*) sounds (0.5–20 kHz broadband, with most energy between 1–10 kHz) from a drifting source boat (Miller et al. 2011, 2012, Curé et al. 2013). From this CEE data set, behavioral responses to MFAS and LFAS have already been scored as minor to moderate by a consensus panel of experts (Miller et al. 2012), and behavioral responses to playback of killer whale sounds have been established and interpreted as antipredatory (Curé et al. 2013). Here, we quantitatively estimated the time and energetic cost of the behavioral response to sonar relative to costs incurred in the antipredatory context as a biologically relevant reference. In order to critically evaluate the baseline variability in sperm whale energy budgets, data from an additional six whales that were tagged in the area but not exposed to sonar or killer whale playbacks were included as baseline data. Potential behavioral changes were also investigated for three of those six whales that were incidentally exposed to 4.75–5.10 kHz sonar signals.

METHODS

Data

Data were collected for 12 sperm whales tagged with audio- and movement-recording data loggers using suction cups (DTAG; Johnson et al. 2009). Four whales were tagged in 2005 (Teloni et al. 2008), and eight whales were tagged in 2008–2010 (Miller et al. 2011, 2012) near the Vesterålen Islands in northern Norway. Sperm whales were located and tracked at sea visually and acoustically by monitoring their echolocation clicks with a towed hydrophone array. The protocol included observations from a main observation vessel at 200

to 1000 m from the whale followed by a launch of a smaller tag boat to deploy the DTAG. Based upon quantified lack of short-term effects after the tagging procedures ended (Isojunno and Miller 2015), the pre-exposure baseline period was defined as the "post-tagging" period that started immediately after the tag boat stopped re-approaching the whale (0.1–2.9 hr after tag-on time, $n = 12$).

Movement sensor data from the tag were decimated at 5 Hz and were used to calculate depth, acceleration, and body pitch angle of the whale (Miller et al. 2011, 2012). Depth data were decimated and pitch data were averaged over 1-min intervals to filter out high-frequency movements such as fluking, but to still allow sufficient time resolution to capture surface intervals. Overall dynamic body acceleration (ODBA, proxy for locomotion effort; Halsey et al. 2009) was calculated as the sum over each minute of the two-norm of high-pass filtered acceleration (symmetric finite impulse response filter, high-pass cut-off frequency 0.05 Hz, length 500 samples). To normalize deployment effects on ODBA such as those due to variable tag position, ODBA values were divided by each whale's median value and then multiplied by the median ODBA across whales. Audio data (stereo, sampled at 96 kHz) were monitored aurally and visually using spectrograms to identify acoustic foraging cues, i.e., echolocation click trains. Rapid increases in click rate (terminal echolocation "buzzes") were used to indicate prey capture attempts (Miller et al. 2004). The presence or absence of prey capture attempts within each 1-min interval was scored using the start time of buzzes.

In order to estimate changes in time allocation to different fitness-enhancing activities, we defined five functional behaviors: (1) surfacing (oxygen replenishment and physiological recovery at the surface); (2) descending transit (transiting to a deeper prey layer); (3) layer restricted search (LRS; searching at a prey layer); (4) ascending transit (transiting to a shallower depth or the surface); and (5) resting and sleep underwater. We also evaluated inclusion of a sixth "other" category of behaviors that could encompass multiple active, but non-foraging functions such as socializing, vigilance, or transiting away from an area. A hidden state-switching model was used to estimate these six behavioral states and associated uncertainty in a Bayesian framework (Isojunno and Miller 2015). The model utilized prior biological information (descent and ascent speed, vertical posture during resting [Miller et al. 2008,], and higher probability of echolocation during foraging), and state-specific likelihoods to describe multiple streams of DTAG data. The model structure, estimation, and resulting time series that were used in the present study are described in Isojunno and Miller (2015). We used output from the lowest deviance information criterion (DIC) model from model selection,

which included state-specific random walk for depth, probability of echolocation (including both regular and terminal buzz clicks), and state-specific relationships between pitch and vertical speed.

Experimental procedures

The experiments were designed and conducted by the 3S (Sea mammals, Sonar, Safety) research project. The full experimental protocol is described in Miller et al. 2011, 2012, and in Curé et al. 2012, 2013, and only briefly summarized here.

Tagged whales were exposed to up to three types of sonar signals: (1) MFAS hyperbolic upsweep at 6–7 kHz, (2) LFAS hyperbolic upsweep at 1–2 kHz, and (3) LFAS hyperbolic downsweep at 1–2 kHz. Source levels were increased over the first 10 min of the exposure ("ramp-up"). The towed source (SOCRATES; Netherlands Organisation for Applied Scientific Research [TNO], The Hague, The Netherlands) was towed toward the whale subject at a depth of about 55 m (range 35–100 m) and source levels (decibel level referenced to one micro Pascal of pressure at 1 meter of distance, dB re 1 µPa m) ranged from 152 to 214 dB for LFAS, and from 158 to 199 dB for MFAS. Sonar signals were 1 s in duration and were transmitted at 20-s intervals. The sonar source was towed, but not transmitting during no-sonar control approaches. The source ship was the 55-m R/V H.U. Sverdrup II.

The experiments could include up to three successive phases. The first phase included 2–3 sonar exposure sessions in which the ship approached the whale transmitting just one sonar signal type or no-sonar negative control. The exposure type (MFAS and LFAS upsweep sonar exposure sessions and no-sonar control) was alternated between sessions. The second phase was dedicated to the playback of killer whale sounds, and the third phase was a final LFAS downsweep sonar exposure session (Miller et al. 2011, 2012). A 1.8–3.2 hr baseline period was recorded after the tag boat left and before the first experimental phase. All exposure and control sessions had at least an hour between them, with 2 hr between the first and second phase. One whale was subject to the first experimental phase only, and three whales were subject to the full protocol. Two tagged whales were subject only to killer whale playback experiments after 8–9 hr of baseline period.

The killer whale sound playback experiments have been published in Curé et al. (2013), and the method is only briefly summarized here. The playback experiments in 2010 included two 15-min sound stimuli broadcast in the following order: a broad band noise playback as negative control and a playback of mammal-eating killer whale sounds. Experiments in 2009 only included the 15-min killer whale sound playback. Killer whale playback stimuli were natural sequences of unfamiliar mammal-eating killer whale sounds previously recorded in the

Northeast Pacific, simulating a potential high risk of predation (Curé et al. 2013). The average total duration of the killer whale vocalizations within each 15 min playback was 7 min 21 s ± 6 s (mean ± SD, $n = 3$ different recordings) (Appendix S4: Fig. S1). The broadband noise control signal was based on non-calling periods during the recordings from which the killer whale sound stimuli were taken, amplified to get an average acoustic power equal to the killer whale sound recordings. Both stimuli were broadcast at a frequency band of 0.5–20 kHz (most energy within 1–10 kHz) at source levels of 145–151 dB re 1 µPa m ($n = 6$ stimuli). The playbacks were conducted from a small motor boat (<10 m) that was stationed at ~800 m from the tagged whale at the start of each playback, and was allowed to drift over the course of the playback (Curé et al. 2013).

Received levels of the sonar signals and source-to-whale range were estimated in Miller et al. (2011). Both the maximum received sound pressure level over a 200-ms window (SPL_{max}; dB re 1 µPa) and the cumulative received sound exposure level (SEL; dB re 1 µPa^2 s) were measured from the tag. SEL was defined as the cumulative sum-of-squared pressures, and measured the cumulative exposure from 1–40 kHz since the start of exposure session (Miller et al. 2011).

Processing of incidental sonar signals

Three whales that were all tagged on 18 July 2005 (sw05_199a, sw05_199b, and sw05_199c) were exposed to incidental sonar produced by an unidentified source at an unknown location. Incidental sonar pings were marked for start and end using the Adobe Audition spectrogram display. For each ping, a time window of at least 200 ms in duration that did not overlap with sperm whale clicks or other noise sources was analyzed. Pings without such a window were marked for start time only, and discarded from received level analysis. Many of the sonar pings contained reverberation, with up to two additional pulse arrivals. In these instances, the first pulse arrival was selected for analysis unless it was masked or was lower in energy than the following one or two pulses, in which case the highest energy pulse was selected.

The analyzed pulse was band-pass filtered (6th-order Butterworth) with frequency cut offs at 4500 and 5400 Hz, and its amplitude corrected for hydrophone sensitivity (−188 dB re 1 µPa re 1V). The sound pressure level averaged over the 90% energy time window $SPL_{90\%}$ in dB re 1 µPa was measured within the filtered signal as

$$SPL_{90\%} = 10\log_{10}\left(\frac{1}{T}\int p^2(t)\, dt\right) \qquad (1)$$

where time window T is defined as the portion of the marked signal window that started at 5% energy and ended at 95% energy ($\int p^2$). The averaging time differed from that for the experimental sonar because the duration of the transmitted incidental sonar signals was <200 ms. The maximum $SPL_{90\%}$ of the two channels was used in the analyses.

Time series modeling of behavioral effects

Three response variables were considered: (1) state (behavioral state); (2) buzz (presence/absence of terminal echolocation clicks [buzz] as a proxy of foraging success); and (3) ODBA (proxy of locomotion activity). State was modeled as a categorical (multinomial) response variable, and presence/absence of buzz as a Bernoulli variable. Presence/absence of buzz was treated for each 1-min interval that reduced the influence of number and duration of buzzes that could potentially vary with prey type. ODBA, as a positive real number, was modeled as a Gamma variable.

A set of baseline explanatory variables was chosen *a priori*. Probability of state, given previous state, was modeled by including previous state (prevState) as factor covariate. State was used as a factor covariate in models for ODBA and buzz in order to allow for mean differences in the proxies for foraging and locomotion across the behavioral budget. Binomial models for buzz were fitted to the subset of data that included foraging states (descent, layer-restricted search [LRS], and ascent). Tag ID (whale) was used as a factor covariate in generalized linear models (GLMs; function multinom in R package nnet for state, and glm function in R package stats for buzz and ODBA; Venables and Ripley 2002, R Development Core 2013) and as a cluster variable in generalized estimating equations (GEE; SAS 9.3 procedure "genmod"; SAS Institute 2011). GLMs and GEEs were used to test which combination of exposure effects best explained the response data (state, buzz, and ODBA). GLMs were used for model selection by Akaike information criterion (AIC), while GEEs were used to allow for serial correlation in the final AIC-selected model to obtain robust individual-average estimates across experiments (Ballinger 2004).

Potential effects of incidental sonar in 2005 (Table 1) were assessed separately from experimental effects due to lack of pre-exposure baseline data for the three incidentally exposed tagged whales. Pre-exposure periods for incidental sonar were not available because, on all three incidentally exposed tags, incidental sonar pings were detected soon after the whales were tagged. Different post-exposure window sizes (10, 15, or 30 min after ping detection; Table 1) and "time since last incidental sonar ping" were assessed as candidate covariates to capture potential recovery to a post-exposure baseline for these whales. Time-decay covariates were calculated as time or time squared, respectively, since last exposure or incidental sonar ping to reflect alternative shapes of recovery from exposure (in a log-linear model, exponential recovery vs. exponential with a delayed speed of recovery).

TABLE 1. List of tested exposure effects on sperm whale behavior (incidental sonar, experimental set-up, and experimental signals of interest) grouped by different scenarios of effect duration (exposure only, exposure + time window, or exposure + time decay).

Covariate type and explanation	Abbreviations		
	Presence/absence	Presence/absence + time window	Time decay
Incidental sonar			
Presence/absence of received pings in the last 10, 15, or 30 min	-	SON05_10, SON05_15, SON05_30	minFromSON05
Maximum $SPL_{90\%}$ in the last 10, 15, or 30 min[†]	-	SPL_max_10, SPL_max_15, SPL_max_30	-
Presence of incidental sonar detection in a given tag	SON05_w	-	-
Exposure session			
Approach of the sonar source vessel	BoatPass	BoatPass_win	minFromBoatPass, min^2FromBoatPass
Order of sonar source vessel approach	#BoatPass	-	-
Playback boat in operation (killer whale sound or noise control playback)	OrcaPB	OrcaPB_win	-
Experimental signal			
Low-frequency active sonar approach	LFAS	LFAS_win	minFromLFAS, min^2FromLFAS
Medium-frequency active sonar approach	MFAS	MFAS_win	minFromMFAS, min^2FromMFAS
Killer whale sound playback (excluding noise playback)	OrcaS	OrcaS	minFromOrcaS, min^2FromOrcaS
Experimental sonar source level	SL	-	-
Experimental sonar received sound pressure level (SPL_{max})	RL	-	-
Experimental sonar cumulative exposure level	SELcum	-	-

Notes: Presence/absence covariate values were set 1 during exposure and 0 otherwise. Presence/absence + time window covariate values were set 1 during exposure, 1 during an estimated window size ("win") min into post-exposure, and 0 otherwise. For experimental effects, the size of time window ("win") was estimated using AIC selection (Appendix S1: Fig. S13).

[†]$SPL_{90\%}$ is the sound pressure level averaged over the 90% energy time window.

The two covariates were set to zero during exposure, and to 900 and 900^2 min for tags that were not exposed to incidental sonar. The 900 min (=15 hr) time was selected to reflect the typical duration of a DTAG record, and to exceed the maximum observed time since last ping for the three incidentally exposed whales (827 min). Since not all tagged whales were exposed to incidental sonar, presence/absence of exposure to incidental sonar over a tag record (SON05_w) was also included as covariate, in case any incidental sonar effect influenced the post-exposure data. No effects of incidental sonar were identified (see section *Results: Test for effects of the incidental sonar exposures*), so data from the post-exposure periods of incidental sonar (30 min after incidental sonar detection) were included as nonexposed data in the following analysis of experimental effects. Model results were tested for robustness to the inclusion of these data (Appendix S2).

Model selection of experimental effects was conducted in two steps. First, any changes in behavior from pre-exposure baseline to during exposure were assessed by excluding all post-exposure data. Any significant effects (exposure types) were retained for the second step, which aimed to estimate the recovery of behavior in the post-exposure period.

For the first step, the candidate covariates included presence/absence of exposure to the three experimental signals (MFAS, LFAS, and OrcaS), presence/absence of exposure to the two types of vessels used to broadcast the signals (BoatPass and OrcaPB), and maximum received sound pressure levels of the sonar signals (SPL_{max}, MFAS:SPL_{max}, LFAS:SPL_{max}) (Table 1). For sonar vessel approaches, an order effect (#BoatPass) allowed any responses to sonar vessel approach to increase or decrease with repetition. The vessel effects (BoatPass, OrcaPB, and #BoatPass) encompassed both signal and control sessions, because the experimental design followed the same protocol in both conditions (an approaching source ship vs. a drifting playback boat). All combinations of the 10 covariates were compared using AIC. The "best" model was selected as the simplest (least parameters) model within two AIC units of the lowest AIC model. If any sonar covariates (MFAS, LFAS, SPL_{max}, MFAS:SPL_{max}, or LFAS:SPL_{max}) were retained in the best AIC model, the selected model was re-fit with source levels and cumulative sound exposure

levels and re-assessed for AIC (Table 1). Again, an AIC decrease of more than two units was considered to improve the model.

In the second step, the selected model was re-fit to all baseline, exposure, and post-exposure data and checked for AIC against the baseline model, in order to assess post-exposure effects. Appropriate window lengths for the different exposure covariates (BoatPass_win, MFAS_win, LFAS_win, OrcaS_win) to encompass any effects that persisted post-exposure were then determined by model selection for each included effect. The model was re-fitted with progressively increasing window size for each effect while keeping other effects fixed. AIC was recorded for every 1-min increase in window size up to 60 min post-exposure. Due to the large number of fitted models (i.e., 60), model averaging was used to select the most appropriate window size. A confidence set of models was selected based upon evidence ratio cut-off ($\exp(-0.5\Delta\text{AIC}) > 0.05$) suggested by Burnham and Anderson (2002). The post-exposure window duration was calculated as the Akaike-weighted mean (Burnham and Anderson 2002) of the minutes in the time window. The newly defined post-exposure window covariates were then compared to models without any post-exposure effects, and with alternative time-decaying post-exposure covariates (Table 1). The simplest (least parameters) model within two AIC units from the lowest AIC model was retained for the final analysis using GEE.

The AIC model selection procedure did not account for serial correlation in the time series, and was therefore likely to retain more effects than could be supported if data were serially correlated. To obtain more robust standard errors, the selected AIC model was refitted within a GEE model, which estimates the empirical correlation within each random effect (whale). All models were fitted within the "genmod" procedure in SAS 9.3 (Littell et al. 2006). The parameter whale was included as a random effect rather than a factor covariate in order to estimate population averaged parameters. The GEE model was fitted using backwards selection where, at each step, one of the coefficients was tested against the null hypothesis that it was zero (Wald test based on empirical standard error estimates) and discarded from the model if the *P* value exceeded 0.05. The procedure was repeated until all remaining explanatory variables were tested $P < 0.05$.

Estimating overall effects on time and energy budgets

The multinomial state models tested for instantaneous changes in Markov transition probabilities, which do not necessarily translate to overall changes in time budgets. Therefore, for those stimuli for which we detected changes in state transitions, a randomization was carried out to test for changes in time budgets given the exposure durations, sampling design, and state uncertainty.

The randomization aimed to test how much time budgets changed from pre-exposure to exposure periods compared to changes observed during baseline tag records. For this analysis, a 90-min pre-exposure baseline period was defined at the start of each baseline record. Short baseline periods ("pseudo-exposures") were then sampled from the subsequent baseline record. Only the five whales with at least 5 hr of baseline data were used to draw these pseudo-exposures (sw05_196a, sw09_153a, sw10_147a, sw10_149a, and sw10_150a).

Pseudo-exposure time budgets were sampled with replacement 10000 times. To account for uncertainty in the state time series, time budgets were sampled from the posterior distribution of states (*n* = 4002 sequences). Each time, a posterior state sequence was drawn from the randomly selected whale, and time budgets were calculated for both the 90-min pre-exposure and pseudo-exposure periods. The pre-exposure time budget was subtracted from the pseudo-exposure budget to give a difference in the proportion of time for each state for each random sample, ranging from −1.0 to +1.0. The resulting distribution of differences (50%, 80%, and 95% quantiles) from pseudo-exposures was then compared to differences in time budgets between actual exposure periods and the 90-min baseline. Actual exposure time budgets were also resampled 10000 times for posterior state sequences. In order to investigate effects of exposure duration on the power to detect changes in time budgets, the resampling procedure was repeated by increasing the duration of the pseudo-exposures from 15 to 90 min at 15-min intervals.

In order to estimate overall changes in the two foraging proxies, and thus indicate overall energy budget, any effects retained in model selection were used to model probability of buzzing (per minute bin) and ODBA excluding state effects in a GEE. Only the relevant exposures, baseline data excluding all post-exposure periods, and the baseline data of whales that were not exposed to incidental sonar were included in the analyses for time and energetic budgets.

Results

Data

A total of 165.5 hr of tag data were analyzed, of which 73.8 hr were pre-exposure baseline data from nine whales (Table 2). Additionally, 43.8 hr of data from three tag records (sw05_199a, sw05_199b, and sw05_199c) that were incidentally exposed to unidentified sonar for 8.7 hr were analyzed. Sonar experiments on four whales (sw08_152a, sw09_141a, sw09_142a, and sw09_160a) included six LFAS (four upsweep and two downsweep) transmission and four MFAS upsweep transmissions, and two no-sonar control approaches

TABLE 2. Duration (hours) of different exposure conditions for each tagged sperm whale (*Physeter macrocephalus*) in the data set.

Tag ID	Total duration (hr)	Pre-exposure baseline (hr)	Sessions with approaching source ship						Playback sessions			
			No-sonar control (hr)		MFAS, 4.7–5.1 kHz or 6–7 kHz (hr)		LFAS, 1–2 kHz (hr)		Noise control (hr)		Killer whale sounds (hr)	
			Dur	Post	Dur	Post	Dur	Post	Dur	Post	Dur	Post
sw05_196a	20.8	20.8	0	0	0	0	0	0	0	0	0	0
sw05_199a	17.6	0	0	0	4.3[†]	13.3[†]	0	0	0	0	0	0
sw05_199b	13.3	0	0	0	3.1[†]	10.2[†]	0	0	0	0	0	0
sw05_199c	12.9	0	0	0	3.6[†]	9.3[†]	0	0	0	0	0	0
sw08_152a	5.8	1.8	0	2.1	1.6	1	1	0.4	0	0	0	0
sw09_141a	14.3	3	1.1	1	0.9	1	0.7	1	0	0	0.4	4.2
sw09_142a	14.5	2.8	0.6	1	0.6	2.5	1.6[‡]	4[‡]	0	0	0.3	1.
sw09_153a	8.4	8.4	0	0	0	0	0	0	0	0	0	0
sw09_160a	14.5	3.2	0	0	0.7	1.7	1.7[‡]	5.3[‡]	0	0	0.3	1.5
sw10_147a	14.8	14.8	0	0	0	0	0	0	0	0	0	0
sw10_149a	14.3	10.1	0	0	0	0	0	0	0.3	0.4	0.3	1
sw10_150a	14.1	8.8	0	0	0	0	0	0	0.3	0.2	0.3	1.2

Notes: "Dur" refers to time periods within the exposure duration, and "Post" refers to post-exposure periods (time period after exposure to next exposure, or end of tag record).
[†] Incidental MFAS exposures.
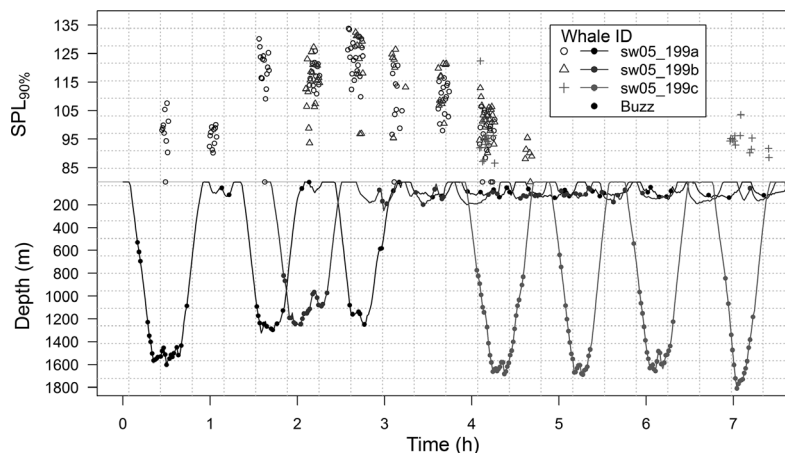[‡] Includes both LFAS upsweep and downsweep exposures.

FIG. 1. Incidental sonar detections and received levels from three tagged sperm whales (*Physeter macrocephalus*). For each tag record, the received sonar levels (sound pressure level averaged over the 90% energy time window [$SPL_{90\%}$], dB re 1 µPa; circles, triangles, and crosses for pings received in each tag record) during incidental sonar exposure, dive profile (1-min down-sampled data), and presence/absence of buzzes (the terminal phase of echolocation used to indicate prey capture attempts; solid circles) used to indicate prey capture attempts; solid circles) are shown. Detected pings for which the received level could not be estimated are plotted on the gray horizontal line (depth = 0).

by the vessel towing the source with no transmission. Received level data for the sonar exposure experiments are reported in Miller et al. (2011). The received median (range) for SPL_{max} was 128.9 (73.2–157.8) dB re 1 µPa for MFAS and 154.0 (72.6–169.6) dB re 1 µPa for LFAS (see Appendix S4: Fig. S42 for sound exposure level and range comparison). Killer whale sound playbacks were conducted on three of the whales that were exposed to sonar experiments and on two additional whales in 2010 (sw10_149a, sw10_150a) that were first exposed to control playbacks of a broadband noise.

Characteristics of the incidental sonar

Incidental sonar pings were detected on tag records from three simultaneously tagged whales. The sonar was received at regular 29.5-s intervals in bouts of 6–22 pings within each tag record (Fig. 1). Each cycle of sonar bouts received in tags sw05_199a ($n = 8$ bouts) and sw05_199b ($n = 6$) consisted of a maximum 10.4 min bout of sonar (mean duration 7.3 min) and a minimum of a 25.0-min interval between consecutive bouts (mean duration 30.8 min). Only two bouts of sonar pings were detected within tag record sw05_199c: 14 pings over a 10.4-min bout in the beginning of the record, and after 2.9 hr, 13 more irregularly spaced pings over a bout of 26.7 min. Sonar pings in the latter bout were also received at similar intervals (minimum of 25.3 s), but this bout had three intermittent silent periods of 3.2, 6.2, and 11.3 min. This was the last bout of sonar pings detected within any of the three tag records.

Incidental sonar pings were 4750–5100 Hz downsweeps of 150-ms duration. Seven out of 222 pings

were removed from analysis due to excessive overlap with sperm whale clicks, and 30 pings had a received $SPL_{90\%}$ below 95 dB re 1 µPa. The median received $SPL_{90\%}$ of the remaining 83% of pings was 113 dB, and the maximum was 134 dB re 1 µPa. All sonar bouts that were detected on tag sw05_199a were also detected on sw05_199b ($n = 5$) and sw05_199c ($n = 1$), except one bout that appeared to end before the whale of sw05_199b left the surface. The received $SPL_{90\%}$ appeared to slowly increase and then decrease over time in all three tags, and the final bout of sonar pings detected in sw05_199c was not detected on either of the other two tags (Fig. 1).

Test for effects of the incidental sonar exposures

There was little apparent difference in time allocation by tagged sperm whales during incidental sonar periods compared to subsequent post-exposure and baseline periods across tag records. During the incidental sonar, the whales performed both very deep (>1000 m) and shallow (~200 m) dives (Fig. 1), and overall, spent a large proportion of time foraging and buzzing (Fig. 2). Within the foraging states, individual average probability of buzzing within 1-min time blocks was higher during incidental sonar exposure in SON05_30 (24.2% ± 10.6% [mean ± SD]) compared to post-exposure baseline (15.6% ± 14.8%; Fig. 2).

There was no statistical support for a change in state transition probability or ODBA within a state during incidental exposures to sonar (Appendix S1: Fig. S12, Table S11). There was weak statistical support for an increase in probability of buzzing during
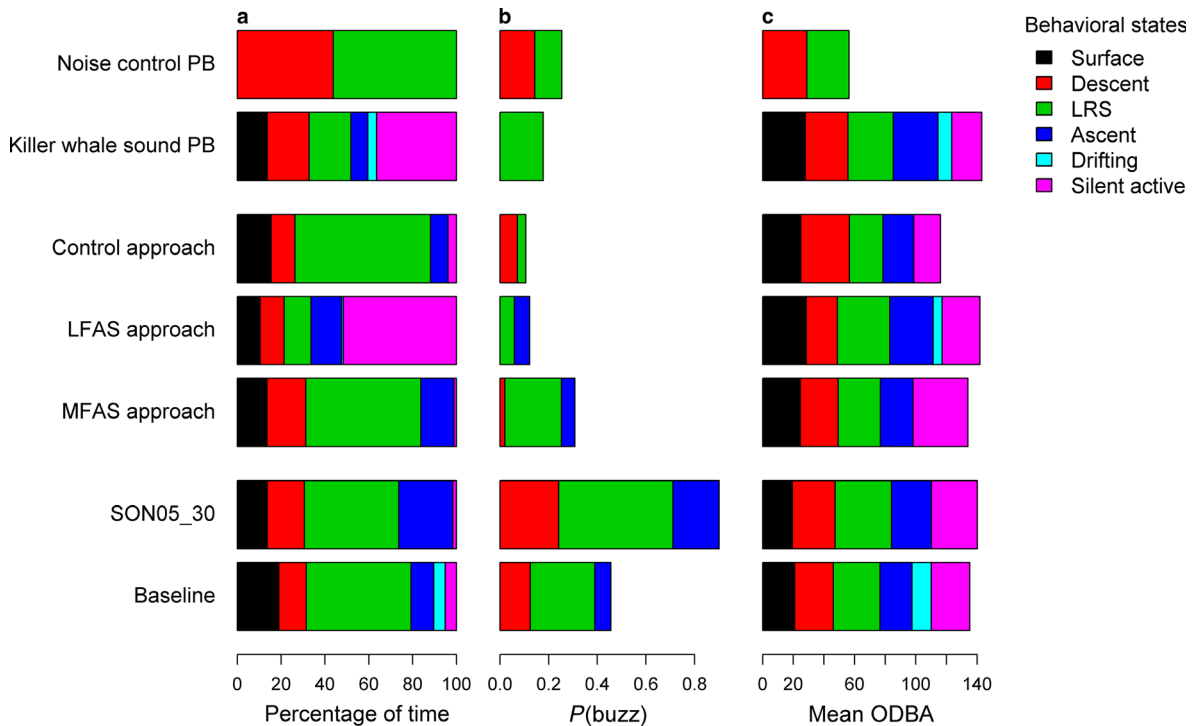
FIG. 2. Time and energy budgets for tagged sperm whales during baseline, incidental sonar (SON05_30, which used a 30-min inter-ping criterion to separate exposure bouts; see Table 2), and experimental exposure sessions (playback, PB): (A) behavioral time budgets; (B) proxies of foraging success (probability of buzzing, as percentage of bins); and (C) locomotion effort (measured as overall dynamic body acceleration; mean ODBA, with higher ODBA values representing increased locomotion effort) averaged across each individual for each state during each experimental control and exposure sessions and baseline (excluding tagging periods and post-exposure periods, but including post-exposure of incidental sonar periods SON05_30). The behavioral states are: (1) surfacing (oxygen replenishment and physiological recovery at the surface); (2) descending transit (transiting to a deeper prey layer); (3) layer restricted search (LRS; searching at a prey layer); (4) ascending transit (transiting to a shallower depth or the surface); (5) resting and sleep underwater; and (6) silent active (multiple active, non-foraging functions such as socializing, vigilance, or transiting away from an area).

incidental sonar exposure in SON05_10 (Wald test, $\chi^2 = 3.61$, $P = 0.057$; Appendix S1: Fig. S12, Table S11). The odds of buzzing were estimated to increase by a factor of 1.74 when pings were detected in the last 10 min (Table 3). However, the effect was weaker for longer post-exposure periods (Appendix S1: Fig. S12, e.g., for SON05_30 ΔAIC < 2) and sw09_199c was highly influential in the model fit (Appendix S2: Fig. S21). The effect of SON05_10, therefore, did not appear robust in the models.

Experimental effects from time series modeling

There was clear statistical support for LFAS and OrcaS as an explanatory variable for state in both AIC model selection (Appendix S1: Tables S12 and S13) and GEE models (Wald tests, *P* values <0.001; Appendix S1: Table S14). LFAS and OrcaS effects were estimated to last until 8 min and 19 min into post-exposure, respectively (Appendix S1: Fig. S13a). Foraging states (descent, layer-restricted search [LRS] or ascent) and active non-foraging state were

estimated as having the greatest magnitude of change during exposures (Appendix S3: Table S31), and were therefore used as the binomial response variables in a GEE. In the binomial GEE for foraging state, the odds of transition to descent, LRS, or ascent was estimated to decrease by a factor of 0.4 both during LFAS_8 and OrcaS_19 (Table 3). Conversely, the odds of transition to an active non-foraging state were estimated to increase by a factor of 6.5 during LFAS exposure until 8 min after exposure (LFAS_8 names the full exposure and 8 min post-exposure period), and increase by a factor 5.7 during OrcaS exposure until 19 min after exposure (OrcaS_19 names the full exposure and 19 min post-exposure period; Table 3).

Buzz presence in the LRS state was less than a quarter of that of the baseline during LFAS and no-sonar approach (6% and 3%, respectively), and somewhat reduced during killer whale sound and noise control playbacks (18% and 11%, respectively) compared to baseline and MFAS (26% and 23%, respectively) (Fig. 2). Model selection for buzz, given a state, retained BoatPass, OrcaPB, and MFAS (Appendix S1:

TABLE 3. Generalized estimating equation (GEE) model estimates.

Model number, response variable, and parameter	Change in odds (e^x)	95% CI (e^x)		Z	P
1) Buzz					
Intercept	0.08	0.05	0.13	−9.9	0.000
Descent	2.61	1.67	4.08	4.2	0.000
LRS	3.91	2.43	6.29	5.6	0.000
SON05_10	1.74	0.98	3.07	1.9	0.057
2) Foraging state (descent, LRS, or ascent)					
Intercept	0.22	0.11	0.44	−4.4	0.000
prevState × surface	0.4	0.2	0.79	−2.6	0.009
prevState × descent	307.2	180.1	523.95	21	0.000
prevState × LRS	612.23	438.61	854.57	37.7	0.000
prevState × ascent	19.71	9.74	39.88	8.3	0.000
prevState × drifting	0.06	0.02	0.24	−4	0.000
LFAS_8	0.41	0.27	0.61	−4.4	0.000
OrcaS_19	0.4	0.21	0.76	−2.8	0.006
3) Non-foraging active state					
Intercept	1.307	0.434	3.942	0.5	0.634
prevState × surface	0.035	0.014	0.086	−7.3	0.000
prevState × descent	0.008	0.004	0.016	−13.3	0.000
prevState × LRS	0.003	0.002	0.005	−20.3	0.000
prevState × ascent	0.003	0.001	0.011	−9.1	0.000
prevState × drifting	0.04	0.013	0.121	−5.7	0.000
LFAS_8	6.513	3.35	12.663	5.5	0.000
OrcaS_19	5.733	3.556	9.242	7.2	0.000
4) Buzz					
Intercept	0.06	0.04	0.09	−12.86	0.000
State × descent	2.78	1.62	4.79	3.69	0.000
State × LRS	5.58	3.74	8.31	8.45	0.000
BoatPass	0.28	0.17	0.46	−5.12	0.000
MFAS	2.69	1.52	4.76	3.41	0.001
OrcaPB	0.6	0.42	0.86	−2.81	0.005

Note: Parameter estimates (back-transformed using exponential function), 95% confidence intervals, and Wald test statistics (*Z* score, *P* value) are shown. "State" represents current behavioral state; "prevState" represents the behavioral state in previous time step (1 min before, given the down-sampling). Numbers following the underscore indicate the duration (min) of the post-exposure time window (e.g., LFAS_8; Table 2).

Tables S12 and S13); no post-exposure effects were supported (Appendix S1: Fig. S13b). Odds of buzzing were estimated to decrease within each state during BoatPass by a factor of 0.28 in a given 1-min bin. Odds of buzzing within each state were also estimated to decrease during OrcaPB, but to a lesser degree (by a factor of 0.60). Relative to LFAS exposures and no-sonar control approaches, odds of buzzing were estimated to be a factor 2.69 higher during MFAS (Table 3).

There were no obvious overall trends in ODBA across states in response to CEE exposures (Fig. 2). Initial model selection for ODBA given a state did retain an effect (Appendix S1: Table S12), but most explanatory variables for ODBA were selected against in the GEE analysis. There was only weak support for a post-exposure recovery after LFAS exposure (minFromLFAS, Wald test, χ^2 = 3.9, *P* = 0.049; Appendix S1: Table S14). ODBA was estimated to

increase slightly as a function of time since LFAS exposure (on average increase of 0.003 for every minute, 95% CI [0.00, 0.006]).

Effects on overall time and energy budget

The individual average percentage of time spent in the active non-foraging state was nearly 10 times higher during LFAS approaches and seven times higher during killer whale sound playbacks compared to baseline periods (Fig. 2). The active non-foraging state was not observed at all during noise control playbacks, and comprised <5.2% of the no-sonar control approach and baseline time budgets. Correspondingly, whales spent on average less than half the time in LRS state during LFAS and killer whale sound playbacks than during baseline, noise control playback, or no-sonar control approach. Whales also spent less time at the surface during LFAS than baseline.

Time budgets during MFAS approach and no-sonar control approach did not differ markedly from baseline (Fig. 2). Because there was little evidence for a change in state transitions or foraging proxies, given a state, in response to MFAS or control signals (see sections *Results: Test for effects of the incidental sonar exposures* and *Experimental effects from time series modeling*), only LFAS and killer whale sound playbacks were considered in the analysis for overall time and energy budgets.

Randomizations within tag records indicated a significant increase in the proportion of time spent in active non-foraging state during most of the LFAS_8 and OrcaS_19 experiments (at two-tailed 5% level, five out of six LFAS_8, and three out of five OrcaS_19 exposures) compared to the 90-min pre-exposure baseline (Fig. 3). Time spent in LRS was more variable, with significant reductions only apparent in sw08_152a (LFAS_8) and sw09_142a (all LFAS_8 and OrcaS_19 exposures). No unusual changes were found for time spent in other states (surface, descent, ascent, or resting; Fig. 3).

Compared to pre-exposure baseline within each whale, an overall decrease in probability of buzzing was observed for all four whales during LFAS (37–100% decrease; two whales ceased buzzing completely during exposure), and four out of five whales exposed to playback of killer whale sounds (45–100% decrease; three whales ceased buzzing completely during exposure). To estimate the effects of LFAS_8 and OrcaS_19 exposures on overall energy budget, probability of buzzing and ODBA were modeled in a GEE excluding state effects. There was good support for an overall reduction in probability of buzzing during LFAS_8 (Wald test, $\chi^2 = 5.05$, $P = 0.025$), but not during OrcaS_19 ($\chi^2 = 0.39$, $P = 0.53$). Overall probability of buzzing was estimated to decrease by a factor of 0.25 (95% CI [0.21, 0.84]) during LFAS_8. There was no support for a change in ODBA during LFAS_8 (Wald test, $\chi^2 = 0.92$, $P = 0.34$), and only weak evidence for a reduction in ODBA during OrcaS_19 ($\chi^2 = 3.33$, $P = 0.068$; Appendix S3: Table S2).

DISCUSSION

We set out to estimate the potential time and energetic cost of behavioral responses of sperm whales to naval sonar. Three statistical hypotheses were tested to assess different possible behavioral responses of sperm whales to naval sonar: (1) changes in
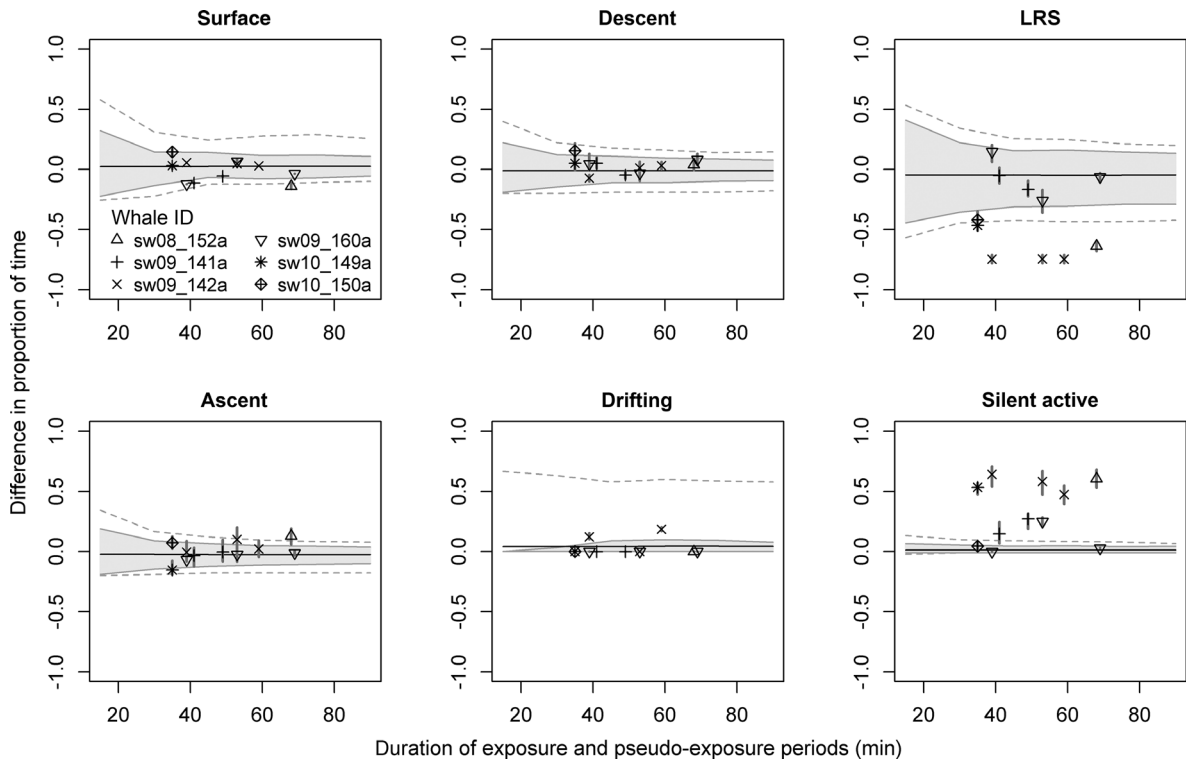

FIG. 3. Differences in the proportion of time spent (*y*-axis) in each behavioral state during exposures and pseudo-exposures compared to the 90-min pre-exposure period. Pseudo-exposures were drawn from short baseline periods that ranged between 15 and 90 min in duration (*x*-axis). Symbols show data during actual killer whale sound playbacks (left, <42 min in duration) and LFAS exposure sessions (right, >49 min in duration), respectively. Dashed lines and shaded areas show 95% and 80% quantiles of the distribution of differences for pseudo-exposure samples, respectively. Vertical gray lines show 95% quantiles for actual exposures.

behavioral state transitions; (2) change in probability of terminal echolocation buzzes (a proxy for foraging success), given behavioral state that could include foraging; and (3) change in overall dynamic body acceleration (ODBA, proxy for locomotion activity), given behavioral state. States were used to capture variability in the two foraging proxies, as we expected both locomotion activity and prey capture attempts to vary both within and across the states, such as lower ODBA during the drifting state. This state-based analysis could therefore differentiate between changes due to behavioral state-switching vs. changes in the proxies within the states, such as potentially reduced prey capture attempts during foraging states due to increased vigilance. Overall, there was clear evidence for a change in functional state transitions to a non-foraging state during LFAS (1–2 kHz sonar) and killer whale sound playbacks, while no such effects were detected for incidental sonar, experimentally presented MFAS, or experimental negative controls (no-sonar approach and noise playback). There was little evidence for changes in locomotion activity (ODBA), while changes in probability of buzzing within states were more uncertain. These results indicate that key behaviors such as foraging were reduced in a trade-off against perceived risk during LFAS approaches and that the level of risk was perceived at a similar level to a mammal-eating predator. While our results imply highly biologically significant changes in behavior, understanding the consequences of such trade-offs requires further research and data on the spatiotemporal overlap of navy sonar with cetacean home ranges, and responsiveness to repeated and longer duration exposures.

Time and energy budgets

Time spent in foraging states and the occurrence of echolocation buzzes was reduced during LFAS and killer whale sound playbacks (Figs. 2 and 3). We addressed correlated time series of state data with a combination of multinomial (categorical) GLM and binomial GEE modeling (see section *Methods: Time series modeling of behavioral effects*). The multinomial GLM models for state appeared effective at detecting and quantifying changes in behavioral state transitions. Although the GLM models assumed that state-transitions only depended upon the previous step (first-order Markov), good concordance was found between predicted and observed time budgets across individuals (Appendix S2: Fig. S22). Binomial GEE models were used to account for any serial dependence beyond the previous step and to ensure significant results were not a mere byproduct of autocorrelation. As statistical tools develop, multinomial GEEs could be used instead. However, both multinomial GLMs and binomial GEE models found significant increases in switching to active non-foraging state, and decreases

in switching to foraging states during LFAS and killer whale sound playbacks (Table 3; Appendix S3: Table S31).

The effect on state transitions was estimated to last for 8 min after the LFAS approach ended and 19 min after the end of killer whale playbacks. Combined with the exposure duration (~30 and 15 min, respectively), this is about the duration of an average sperm whale dive or a short dive cycle (40 min; Watwood et al. 2006). Similarly, randomization of time budgets from individual baseline data indicated that changes in state transitions translated to changes in time budgets both during LFAS and killer whale sound playbacks at time scales exceeding ~40 min (Fig. 3). These results suggest a functional shift in behavior from foraging to non-foraging, which is also supported by expert scoring of cessation of feeding, change in vocal behavior, and change in dive behavior during the same LFAS and killer whale sound exposures, and lack thereof during MFAS and no-sonar control approaches (Miller et al. 2012, Curé et al. 2013). Similarly, Sivle et al. (2012) reported more consistent changes in diving (shallower deep dives) during LFAS than MFAS for these whales. Whales did not switch to resting or increase resting time during exposures, indicating that reduced energy intake was not immediately compensated for, as was suggested by Miller et al. (2009) for sperm whales exposed to seismic signals.

There is an increasing appreciation that both acoustic and behavioral context modulates responsiveness to disturbance (Ellison et al. 2012). In our sample, four out of six exposed whales reduced time spent in layer-restricted search (LRS) by 66–100% during LFAS approaches and killer whale sound playbacks compared to pre-exposure baseline. The response profiles of the two other whales (sw09_141 and sw09_160) were not as clear. Whale sw09_141 appeared to be engaging in social interactions in a group of two to four whales during pre-exposure baseline, with 46% of time spent in active non-foraging state. Therefore, sw09_141 was already mostly in an active non-foraging state during the pre-exposure period and remained mainly in this state throughout LFAS and killer whale playbacks. In contrast, sw09_160 spent 87% of the baseline period in foraging states (descent, LRS, and ascent) and had the highest baseline probability of buzz occurrence in LRS state (58% of 1-min duration analysis blocks) than any other whales in this study. Whale sw09_160 continued foraging during most of the exposures to both LFAS and killer whale sounds. Although anecdotal as evidence, these two whales demonstrate two contrasting behavioral contexts (socializing vs. intense foraging) that could be important in modulating their responsiveness to anthropogenic stimuli, as well as our ability to detect change in foraging behavior. Further studies utilizing larger sample sizes would be needed to verify whether sperm whales are more responsive

during socializing or foraging. For example, such behavioral context-specific effects were found for northern and southern resident killer whales (Williams et al. 2006, Lusseau et al. 2009). Besides identifying responsive behavior states, studies monitoring pre-exposure behavioral context could also help to estimate effect sizes of future CEEs that inherently have relatively small sample sizes.

Foraging proxies

Buzz presence was reduced overall in response to LFAS because the animals transitioned out of the foraging states, but within a foraging state, there was no clear support for a change in probability of buzzing during any experimental signal exposure. However, buzzing only occurred during the foraging states (descent, LRS, and ascent), and the reduced foraging time observed during LFAS and killer whale sound playbacks subsequently reduced the power of the statistical test to detect significant effects on buzzing, given a state (Fig. 3). Probability of a buzz during foraging states was estimated to decrease during all approaches, with or without sonar, but there was less effect during MFAS approaches (Table 3; Appendix S2: Fig. S27). While one of the no-sonar control approaches was conducted before the full CEE (sw09_142a; LFAS first, MFAS second), the other control approach was conducted after the LFAS and MFAS exposures (sw09_141a). With a brief transition to active non-foraging state during the ascent phase in the latter control, it is possible that the whale (sw09_141a) had become sensitized to the source vessel. Indeed, the contribution of BoatPass to the GLM model AIC was almost twice as much as the contribution of MFAS (Appendix S1: Table S13). Future analyses contrasting multiple exposures and controls, and thus larger number of model structures, could reduce uncertainty in model selection by averaging predictions from a "confidence set" of models (Burnham and Anderson 2002, see Isojunno et al. 2012 for an application including temporally varying covariates).

While the GEE analysis excluding state effects suggested an overall reduction in probability of buzzing during LFAS, there was no evidence for an overall change in probability of buzzing during killer whale sound playbacks (Appendix S3: Table S32). These results indicate that, although an overall individual-average reduced probability of buzzing could be detected for LFAS, the short-term reductions in probability of buzzing during killer whale playbacks (that were short compared to LFAS exposures) could have been achieved by chance alone if reduction of time in foraging states were not considered. This was expected as sperm whales spend time in functionally important non-foraging behaviors (surfacing, resting, socializing), and therefore short cessations of feeding are not unusual.

There was weak evidence for an increase in the probability of prey capture attempts (buzzes) during the detected bouts of incidental sonar compared to post-exposure and baseline data across tag records (Table 3). However, it is likely that the marginally supported effect size was artificially inflated by confounding effects of dive depth on incidental sonar detection, as well as the potential for pseudo-replication of similar foraging conditions. In other words, the three whales tagged in the same day were not independent samples of a behavioral response to the sonar, and therefore local conditions such prey type and availability may have influenced the statistical result.

There was little evidence for change in locomotion activity, given a state, during incidental sonar or experimental exposures (Fig. 2). The small decrease in overall ODBA during killer whale sound playbacks may have reflected a reduced foraging effort. The distribution of ODBA values during non-foraging active state was nearly identical to that of three foraging states during baseline and therefore switching between these states alone was unlikely to result in changes in overall ODBA. Marine mammals are efficient swimmers that have evolved to minimize cost of transport at a range of swimming speeds (Williams 1999). It therefore seems likely that whales switched behavioral mode with consequences on energy intake, rather than expenditure on locomotion.

Comparison of sonar stimuli

In contrast to LFAS, no reduction in foraging effort was detected during no-sonar approaches, experimental MFAS, or incidental sonar exposures (Figs. 1 and 2). From the same data, expert judgment also scored higher severity responses to LFAS than MFAS for sperm whales (Miller et al. (2012). These results indicate that sperm whales mitigated a higher perceived risk or cost during LFAS approaches compared to other sonar exposures. The differential response may have been due to higher source level, lower frequency, different sensation level, and/or unfamiliarity of the experimental LFAS exposures.

LFAS was both transmitted and received at higher sound pressure levels (received SPL_{max} 170 dB re 1 μPa) than MFAS (received SPL_{max} 158 dB re 1 μPa). Incidental sonar pings were generally received at the lowest sound pressure levels (max $SPL_{90\%}$ 133.5 dB re 1 μPa). The temporal patterns in received $SPL_{90\%}$ were surprisingly consistent across the three incidentally exposed tags (Fig. 1), indicating that the movement of the whales had relatively minor influence on the received levels. For this reason, and because the relative levels of reverberation were much higher than for the experimental sonar, it seems reasonable to assume that the incidental sonar was relatively distant (tens of kilometers). Naval sonars may be detectible at ranges up to tens to hundreds of kilometers (Hildebrand 2009).

However, our observations appear to have started after the incidental sonar transmissions were underway. Therefore, our data indicating low responsiveness to incidental sonar does not allow any conclusion on whether an initial response to the incidental sonar could have been similarly strong as to the experimental LFAS.

The frequency of the incidental sonar (4.75–5.10 kHz downsweep) was closer to the fundamental frequency of MFAS upsweeps (6–7 kHz) than LFAS upsweeps or downsweeps (1–2 kHz). The LFAS sonar source produced more harmonic distortions when transmitting at higher source levels (Miller et al. 2011), although individual harmonics were less intense (<-15 dB) than the fundamental, with little difference between the broadband received SPL_{max} and the SPL_{max} within the 1–2 kHz operational band. An operational sonar source was used in these exposure experiments, and since operational naval sonars typically have harmonic distortions at full power, this was part of the realism of the experimental design. The low duty cycle of sonar reduces concern about masking, but it is possible that the received sonar, including associated reverberation and harmonics, may have potential to mask portions of returning echolocation signals that contain information about prey.

While there was no evidence of response for the relatively low received-level incidental sonars (Fig. 1), there were clear changes in behavior of the subjects during LFAS experimental exposures. However, the state transition analysis did not lend support for a particular response threshold within the experimentally tested range of received MFAS or LFAS. Presence/absence of LFAS was retained in model selection over main effects or interactions of sonar presence with received levels. These interactions allowed the effect of received level to vary between the two frequency bands. The analysis only used unweighted broadband levels due to lack of data on hearing sensitivity for sperm whales. M-weighting was not applied because all signal frequencies fell within the flat passband of the M-weighting filter response specified for sperm whales by Southall et al. (2007). In future studies on species for which auditory weighting functions are available, and appropriate to use given the signal frequencies, testing weighted vs. unweighted received level as explanatory variables may be a powerful means to test which type of weighting function best predict behavioral response thresholds.

All four exposed whales appeared to start switching to active non-foraging state at a cumulative SEL of 135–145 dB, which was ~20 dB below the maximum cumulative SEL of MFAS (median 139.6 dB, max 160.6 dB). If the sperm whale hearing filter tapered off below 2–5 kHz (Ridgway and Carder 2001), a lower detection threshold would be expected for the MFAS relative to the LFAS fundamental frequency. On the other hand, a significantly lower sensitivity at 1–2 kHz would appear somewhat disadvantageous if

lower frequencies were also important for long-range detection of conspecifics or mammal-eating killer whales. Switching to active non-foraging state was observed within minutes of onset of LFAS exposure and both in final ascent and first descent of foraging dives (sw08_152a and sw09_142a), as well as in surface state (sw09_160a). These responses suggest that the received sound pressure levels of LFAS did not necessarily interfere with foraging (e.g., masking of echolocation, effects on prey), and that perceptual effects (signal interpretation, experience) may have been important in how sperm whales responded to sonar signals.

Risk–disturbance hypothesis

The risk–disturbance hypothesis predicts trade-offs between investment in critical activities such as foraging and safety from predators and anthropogenic threats. Our time series models support this hypothesis, showing nearly identical decreases in the probability of state transition to foraging between vessel approaches transmitting LFAS and playback of predator (killer whale) sounds (odds decreased by a factor of 0.4). Assuming that time and energy are traded off as a function of perceived risk, sperm whales perceived an LFAS approach (from 10 km to a minimum approach distance of 320 m) as high a risk as detection of a potential predator at ranges <1 km. More direct and fast approaches of threats have been shown to increase perceived risk and subsequent responsiveness in a range of species (Frid and Dill 2002, Stankowich and Blumstein 2005, Domenici et al. 2011). Similarly here, experimental LFAS and MFAS were transmitted on a controlled distance approach to focal whales, and may have triggered a stronger response (reduction in buzzing) as a result compared to the near-stationary playback, or to incidental sonar, which did not perform a close approach.

The perceived acoustic signature of LFAS may have been associated with increased risk by resembling a specific predator-like acoustic signal (Sih 2013), and/or its features may be less specific indicators of threat, such as low fundamental frequency indicating a larger source (Fitch 2000). The killer whale is a cosmopolitan species with localized ecotypes (de Bruyn et al. 2013), and it is possible that sperm whales are wary of any sound resembling their vocalizations, e.g., due to matching spectral characteristics. From a predator–prey arms race perspective, signals that are less identifiable or more cryptic at the edge of audibility may cue for increased risk as a potential indicator of predator intent to avoid detection by prey. However, it seems likely that sperm whales could differentiate between sonar transmission and killer whale sounds. Unlike motor vessels, which produce sound continuously while underway, killer whales can move silently. Mammal-eating killer whales can cease vocalizing

before engaging in a cryptic hunting strategy (Deecke et al. 2005). Thus, the ceasing of predator vocalizations could be a signal that increased the perceived predation risk and subsequent antipredatory response in the post-exposure period of the predator sound playbacks. Indeed, our analysis indicated a longer post-exposure effect for the killer whale sound playback compared to LFAS.

CONCLUSIONS

Behavioral state transition modeling showed that tagged whales switched to a non-foraging, active state during both experimental transmissions of low-frequency active sonar (LFAS; 1–2 kHz) and playbacks of mammal-eating killer whale vocalizations. Time spent in foraging states and prey capture attempts were reduced during these two types of exposures, with little change in overall locomotion activity, suggesting an effect on energy intake rather than on expenditure. In contrast, no changes in foraging behavior were detected during experimental controls (no-sonar approach or near-stationary noise playback) or experimental medium-frequency active sonar exposures (MFAS; 6–7 kHz) or exposure to unidentified 4.7–5.1 kHz sonar signals. Given the relatively short duration of the ramp-up and small sample size, there was little power to test effects of spectral signature and source level as separate disturbance stimuli. Nevertheless, sensitivity to higher source levels or the specific 1–2 kHz frequency band with harmonics remain plausible factors explaining why sperm whales were more disturbed by 1–2 kHz LFAS sonar used in this study than the 6–7 kHz MFAS sonar or the 4.7–5.1 kHz incidental sonar signals.

Our small sample size of individuals (12 tags, 6 experimentally exposed) reflected the expense and logistics of conducting CEEs while requiring sufficient data to quantify potential changes from "natural" or baseline behavior, which alone can be highly variable between contexts and individuals. The statistical analysis incorporating foraging context helped to control inter-individual variability and thus reveal significant changes in behavior, but the extrapolation of our results to larger temporal and spatial scales, including populations, is inherently limited. Long-term monitoring of individuals (e.g., satellite tagging with "smart" programming of daily time budget), coarser temporal sampling of more individuals (e.g., land-based theodolite tracking; Williams et al. 2014) and acoustic monitoring of both individuals and populations (e.g., Castellote et al. 2012, Moretti et al. 2014) before, during, and after experimental or opportunistic exposures could be used to validate the predictions generated from our CEE, in particular: (1) low response thresholds of sperm whales to low-frequency impulsive sounds, and (2) habituation to or noise tolerance of medium-frequency impulsive sounds.

An advantage of state-based approaches in behavioral response studies is that they provide an explicit prediction of how changes in behavior might be extrapolated for a given exposure duration. Activity budgets can be defined at multiple scales, and tuned to particular study systems depending upon the behavioral spectrum of the species and available data. For example, odontocete foraging may be monitored at fine temporal scales using passive acoustic monitoring, defecation rates at medium scales, and movement between foraging and breeding grounds at much coarser scales. Changes in time and energy budgets can therefore more directly provide data for individual-based models that aim to propagate short-term and individual effects to long-term and population level, such the Population Consequences of Acoustic Disturbance (NRC [National Research Council] 2005). Our results indicate that the overall reduction in probability of a prey capture attempt (in 1-min time blocks) and increased time spent in active non-foraging modes of behavior in sperm whales could lead to a significant reduction in prey intake if the behavioral changes persisted over longer or repeated exposures to either LFAS or predator sounds. This prediction is similar to the results of Williams et al. (2006), who, by assigning fixed energy expenditures to visually classified activity states, found that altered time allocation in killer whale groups exposed to vessel traffic resulted in a reduced energy intake rather than increased energy expenditure. We demonstrated that behavioral responses to anthropogenic noise and predator stimuli induced similar reduction in foraging time in sperm whales, suggesting similar levels of perceived threat. Nevertheless, the temporal extent to which sperm whales are exposed to types of naval sonar at regional or global scales is not known, nor whether habituation/sensitization processes would allow for subsequent chronic reductions in foraging success. Behavioral responses to predators and generalized threatening stimuli are likely based upon similar response mechanisms, and such behaviors may have been selected in response to signals whose recognition ranges from uncertain to highly certain (Sih 2013). Future research on sperm whale behavioral responses could address whether specific antipredatory behaviors such as acoustic crypsis or a social response (Curé et al. 2013) are also used in the context of anthropogenic disturbance stimuli.

Literature Cited

Alves, A., R. Antunes, and A. Bird. 2014. Vocal matching of naval sonar signals by long-finned pilot whales (*Globicephala melas*). Marine Mammal Science 30:1248–1257.

Antunes, R., P. H. Kvadsheim, F. P. A. Lam, P. L. Tyack, L. Thomas, P. J. Wensveen, and P. J. O. Miller. 2014. High thresholds for avoidance of sonar by free-ranging long-finned pilot whales (*Globicephala melas*). Marine pollution bulletin 83:165–180.

Balcomb, K. C. III, and D. E. Claridge. 2001. A mass stranding of cetaceans caused by naval sonar in the Bahamas. Bahamas Journal of Science 8:2–12.

Ballinger, G. A. 2004. Using generalized estimating equations for longitudinal data analysis. Organizational Research Methods 7:127–150.

Beale, C. M. 2007. The behavioral ecology of disturbance responses. International Journal of Comparative Psychology 20:111–120.

Bejder, L., A. Samuels, H. Whitehead, H. Finn, and S. Allen. 2009. Impact assessment research: use and misuse of habituation, sensitisation and tolerance in describing wildlife responses to anthropogenic stimuli. Marine Ecology Progress Series 395:177–185.

Berger-Tal, O., T. Polak, A. Oron, and Y. Lubin. 2011. Integrating animal behavior and conservation biology: a conceptual framework. Behavioral Ecology 22:236–239.

Best, P. B. 1979. Social organization in sperm whales, *Physeter macrocephalus*. Plenum Press, New York, New York, USA.

Brown, J. S., and B. P. Kotler. 2004. Hazardous duty pay and the foraging cost of predation. Ecology Letters 7:999–1014.

de Bruyn, P. J. N., C. A. Tosh, and A. Terauds. 2013. Killer whale ecotypes: is there a global model? Biological reviews of the Cambridge Philosophical Society 88:62–80.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York, New York, USA.

Castellote, M., C. W. Clark, M. O. Lammers. 2012. Acoustic and behavioural changes by fin whales (*Balaenoptera physalus*) in response to shipping and airgun noise. Biological Conservation 147:115–122.

Clark, C. W., W. T. Ellison, B. L. Southall, L. Hatch, S. M. Van Parijs, A. Frankel, and D. Ponirakis. 2009. Acoustic masking in marine ecosystems: intuitions, analysis, and implication. Marine Ecology Progress Series 395:201–222.

Cox, T. M., et al. 2006. Understanding the impacts of anthropogenic sound on beaked whales. Journal of Cetacean Research and Management 7:177–187.

Curé, C., R. Antunes, F. Samarra, A. C. Alves, F. Visser, P. H. Kvasheim, and P. J. O. Miller. 2012. Pilot whales attracted to killer whales sounds: acoustically-mediated interspecific interactions in cetaceans. PLoS ONE 7:e52201.

Curé, C., R. Antunes, A. C. Alves, F. Visser, and P. H. Kvadsheim. 2013. Responses of male sperm whales (*Physeter macrocephalus*) to killer whale sounds: implications for anti-predator strategies. Scientific Reports 3:1–7.

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

Deecke, V., J. Ford, and P. Slater. 2005. The vocal behaviour of mammal-eating killer whales: communicating with costly calls. Animal Behaviour 69:395–405.

DeRuiter, S. L., I. L. Boyd, D. E. Claridge, C. W. Clark, C. Gagnon, B. L. Southall, and P. L. Tyack. 2013a. Delphinid whistle production and call matching during playback of simulated military sonar. Marine Mammal Science 29:46–59.

DeRuiter, S. L., et al. 2013b. First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. Biology Letters 9:1–5.

Domenici, P., J. M. Blagburn, and J. P. Bacon. 2011. Animal escapology I: theoretical issues and emerging trends in escape trajectories. Journal of Experimental Biology 214:2463–2473.

Ellison, W. T., B. L. Southall, C. W. Clark, and A. S. Frankel. 2012. A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds. Conservation Biology 26:21–28.

Engelhaupt, D., et al. 2009. Female philopatry in coastal basins and male dispersion across the North Atlantic in a highly mobile marine species, the sperm whale (*Physeter macrocephalus*). Molecular Ecology 18:4193–4205.

Fahlman, A., P. L. Tyack, P. J. Miller, and P. H. Kvadsheim. 2014. How man-made interference might cause gas bubble emboli in deep diving whales. Frontiers in Physiology 5:1–6.

Filadelfo, R., J. Mintz, E. Michlovich, A. D. Amico, P. L. Tyack, and D. R. Ketten. 2009. Correlating military sonar use with beaked whale mass strandings: what do the historical data show? Aquatic Mammals 35:435–444.

Fitch, W. T. 2000. Skull dimensions in relation to body size in nonhuman mammals: The causal bases for acoustic allometry. Zoology-Analysis of Complex Systems 103:40–58.

Frantzis, A. 1998. Does acoustic testing strand whales? Nature 392:29.

Frid, A., and L. Dill. 2002. Human-caused disturbance stimuli as a form of predation risk. Conservation Ecology 6:11.

Gill, J. A., K. Norris, and W. J. Sutherland. 2001. Why behavioural responses may not reflect the population consequences of human disturbance. Biological Conservation 97:265–268.

Goldbogen, J. A., et al. 2013. Blue whales respond to simulated mid-frequency military sonar. Proceedings of the Royal Society B 280:1–8.

Halsey, L. G., E. L. C. Shepard, F. Quintana, A. Gomez Laich, J. A. Green, and R. P. Wilson. 2009. The relationship between oxygen consumption and body acceleration in a range of species. Comparative Biochemistry and Physiology A 152:197–202.

Hildebrand, J. A. 2009. Anthropogenic and natural sources of ambient noise in the ocean. Marine Ecology Progress Series 395:5–20.

Isojunno, S., and P. J. O. Miller. 2015. Sperm whale response to tag boat presence: biologically informed hidden state models quantify lost feeding opportunities. Ecosphere 6:1–46.

Isojunno, S., J. Matthiopoulos, and P. H. Evans. 2012. Harbour porpoise habitat preferences: robust spatio-temporal inferences from opportunistic data. Marine Ecology Progress Series 448:155–170.

Johnson, M., N. Aguilar de Soto, and P. T. Madsen. 2009. Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags. Marine Ecology Progress Series 395:55–73.

Kvadsheim, P. H., P. J. O. Miller, P. Tyack, L. D. Sivle, F. P. A. Lam, and A. Fahlman. 2012. Estimated tissue

and blood N_2 levels and risk of *in vivo* bubble formation in deep-, intermediate- and shallow-diving toothed whales during exposure to naval sonar. Frontiers in Physiology 3:125.

Littell, R., G. Milliken, W. Stroup, R. Wolfinger, and O. Schabenberger. 2006. SAS for mixed models. SAS Institute, Cary, North Carolina, USA.

Lusseau, D., D. E. Bain, R. Williams, and J. C. Smith. 2009. Vessel traffic disrupts foraging behaviour of southern resident killer whales. Endangered Species Research 6:211–221.

Miller, P. J. O., M. P. Johnson, and P. L. Tyack. 2004. Sperm whale behaviour indicates the use of echolocation click buzzes "creaks "in prey capture. Proceeding of the Royal Society of London Series B 217:2239–2247.

Miller, P. J. O., K. Aoki, L. E. Rendell, and M. Amano. 2008. Stereotypical resting behavior of the sperm whale. Current Biology 18:R21–R23.

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

Miller, P., et al. 2011. The 3S experiments: studying the behavioural effects of naval sonar on killer whales (*Orcinus orca*), sperm whales (*Physeter macrocephalus*), and long-finned pilot whales (*Globicephala melas*) in Norwegian waters. Technical Report SOI-2011-001. Scottish Oceans Institute, St Andrews, Fife, UK.

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

Miller, P. J. O., et al. 2014. Dose-response relationships for the onset of avoidance of sonar by free-ranging killer whales. Journal of the Acoustical Society of America 35:975–993.

Moretti, D., et al. 2014. A risk function for behavioral disruption of Blainville's beaked whales (*Mesoplodon densirostris*) from mid-frequency active sonar. PLoS ONE 9:e85064.

NRC [National Research Council]. 2005. Marine mammal populations and ocean noise: Determining when noise causes biologically significant effects. US National Academy of Sciences, Washington, D.C., USA.

Oliveira, C., and M. Wahlberg. 2013. The function of male sperm whale slow clicks in a high latitude habitat: communication, echolocation, or prey debilitation? Journal of the Acoustical Society of America 133:3135–3144.

de Quirós, Y. B., O. González-Diaz, M. Arbelo, E. Sierra, S. Sacchini, and A. Fernández. 2012. Decompression vs. decomposition: distribution, amount, and gas composition of bubbles in stranded marine mammals. Frontiers in Physiology 3:1–19.

R Development Core, T.. 2013. R: a language and environment for statistical computing Release 3.0.1. R Foundation for Statistical Computing, Vienna, Austria.

Reeves, R. 2003. Dolphins, whales and porpoises: 2002–2010 conservation action plan for the world's cetaceans. IUCN, Gland, Switzerland.

Ridgway, S., and D. Carder. 2001. Assessing hearing and sound production in cetaceans not available for behavioral audiograms: Experiences with sperm, pygmy sperm, and gray whales. Aquatic Mammals 27:267–276.

SAS Institute. 2011. The SAS system for Windows Release 9.3. Cary, North Carolina USA, SAS Institute.

Sih, A. 2013. Understanding variation in behavioural responses to human-induced rapid environmental change: a conceptual overview. Animal Behaviour 85:1077–1088.

Sivle, L. D., P. H. Kvadsheim, A. Fahlman, F. P. A. Lam, P. L. Tyack, and P. J. O. Miller. 2012. Changes in dive behavior during naval sonar exposure in killer whales, long-finned pilot whales, and sperm whales. Frontiers in Physiology 3:400.

Southall, B. L., et al. 2007. Marine mammal noise-exposure criteria: initial scientific recommendations. Aquatic Mammals 33:411–521.

Stankowich, T., and D. Blumstein. 2005. Fear in animals: a meta-analysis and review of risk assessment. Proceedings of the Royal Society B 272:2627–2634.

Teloni, V., J. P. Mark, M. J. O. Patrick, and M. T. Peter. 2008. Shallow food for deep divers: dynamic foraging behavior of male sperm whales in a high latitude habitat. Journal of Experimental Marine Biology and Ecology 354:119–131.

Tyack, P. 2009. Acoustic playback experiments to study behavioral responses of free-ranging marine animals to anthropogenic sound. Marine Ecology Progress Series 395:187–200.

Tyack, P. L., et al. 2011. Beaked whales respond to simulated and actual navy sonar. PLoS ONE 6:e17009.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S, Fourth edition. Springer, New York, New York, USA.

Watwood, S. L., P. J. O. Miller, M. Johnson, P. T. Madsen, and P. L. Tyack. 2006. Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*). Journal of Animal Ecology 75:814–825.

Williams, T. M. 1999. The evolution of cost efficient swimming in marine mammals: limits to energetic optimization. Proceedings of the Royal Society B 354:193–201.

Williams, R., D. Lusseau, and P. S. Hammond. 2006. Estimating relative energetic costs of human disturbance to killer whales (*Orcinus orca*). Biological Conservation 133:301–311.

Williams, R., C. Erbe, E. Ashe, A. Beerman, and J. Smith. 2014. Severity of killer whale behavioral responses to ship noise: a dose-response study. Marine Pollution Bulletin 79:254–260.

Wright, A. J., T. Deak, and E. C. M. Parsons. 2011. Size matters: management of stress responses and chronic stress in beaked whales and other marine mammals may require larger exclusion zones. Marine Pollution Bulletin 63:5–9.

Supporting Information

Data availability