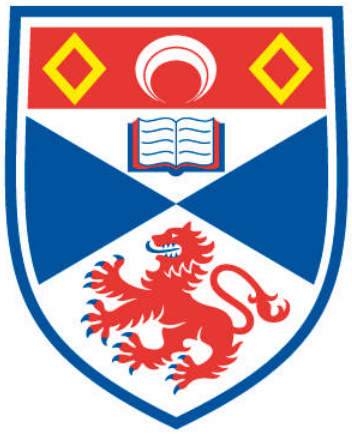
DETECTING, ASSESSING, AND MITIGATING THE EFFECTS OF NAVAL SONAR ON CETACEANS

Paul J. Wensveen

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2016

Detecting, Assessing, and Mitigating the Effects of Naval Sonar on Cetaceans

Paul J. Wensveen

University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of PhD
at the
University of St Andrews

26 December 2015

The truth is: the natural world is changing. And we are totally dependent on that world.

– Sir David Attenborough

Summary

Effective management of the potential environmental impacts of naval sonar requires quantitative data on the behaviour and hearing physiology of cetaceans. Here, novel experimental and analytical methods were used to obtain such information and to test the effectiveness of an operational mitigation method for naval sonar. A Bayesian method was developed to estimate whale locations through time, integrating visual observations with measurements from on-animal inertial, acoustic, depth, and Fastloc-GPS sensors. The track reconstruction method was applied to 13 humpback whale (*Megaptera novaeangliae*) data sets collected during a multi-disciplinary behavioural response study in Norwegian waters. Thirty-one controlled exposure experiments with and without active transmissions of 1.3-2 kHz sounds were conducted using a moving vessel that towed a sonar source. Dose-response functions, representing the relationships between measured sonar dose and behavioural responses identified from the reconstructed tracks, predicted that 50% of the humpbacks would initiate avoidance at a relatively high received sound pressure level of 166 dB re 1 μPa. Very similar dose-response functions were obtained for cessation of feeding. In a laboratory study, behavioural reaction times of a harbour porpoise (*Phocoena phocoena*) to sonar-like sounds were measured using operant conditioning and a psychoacoustic method. Auditory weighting functions, which can be used to improve dose-response functions, were obtained for the porpoise based on the assumption that sounds of equal loudness elicit equal reaction time. Additional analyses of the humpback whale data set provided evidence that ramp-up of naval sonar mitigates harmful sound levels in responsive cetaceans located directly in the path of the source, and suggested that a subset of the humpback whale population, such as mother-calf pairs, and more responsive species would benefit from the use of sonar ramp-up. The findings in this thesis are intended to inform sound exposure criteria and mitigation guidelines for anthropogenic noise exposure to cetaceans.

Declarations

I, Paul Wensveen, hereby certify that this thesis, which is approximately 51,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September, 2011 and as a candidate for the degree of PhD in September, 2011; the higher study for which this is a record was carried out in the University of St Andrews between 2011 and 2015.

Date: Signature of candidate:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: Signature of supervisor:

Permission for Publication

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisors regarding the publication of this thesis:

No embargo on print copy and electronic copy

Date: Signature of candidate:

Date: Signature of supervisor:

Collaboration Statement

Collection of the humpback whale data used in Chapters 2, 3 and 5 was carried out in a multi-disciplinary project ("3S Behavioural Response Study"). The data collection and tagging protocols had already been established, but I was involved in designing the experimental protocol for the ramp-up study (Chapter 5). I conducted focal follows and helped with the navigation of the source vessel during field work in 2011 and 2012.

I designed the position accuracy tests (Chapter 2) in collaboration with my principal supervisor Patrick Miller and conducted the tests. My secondary supervisor Len Thomas helped me with the development of the track reconstruction model in Chapter 2. Len Thomas and Catriona Harris advised me on the dose-response analysis conducted in Chapter 3. I conducted part of the acoustic analyses of Chapters 3 and 5 at TNO in The Hague in collaboration with Sander von Benda-Beckmann (noise measurements with DTAG), Michael Ainslie (acoustic propagation modelling) and Wim Groen (DTAG calibrations). The expert scoring of the behavioural responses in Chapter 3 was in collaboration with Fleur Visser, Charlotte Curé and Marjoleine Roos. Detection of the feeding lunges used in these chapters was conducted by Lise Sivle.

I worked together with Ron Kastelein to gain funding and design the experimental protocol for the porpoise study (Chapter 4). The reaction time data was collected and processed by Lean Hoek, Leonie Huijser and myself. The reaction time sensor was designed by Arie Smink. The acoustic calibrations in Chapter 4 were conducted by Erwin Jansen. I conducted all further data analyses.

Parts of Chapters 2 and 4 of this thesis were published in peer-reviewed journals:

- Wensveen, P.J., Thomas, L., and Miller, P.J.O. (2015). A path reconstruction method integrating dead-reckoning and position fixes applied to humpback whales. Movement Ecology 3, 31.
- Wensveen, P.J., Huijser, L.A.E., Hoek, L., and Kastelein, R.A. (2015). Equal latency contours and auditory weighting functions for the harbour porpoise (*Phocoena phocoena*). Journal of Experimental Biology 217, 359-369.

Acknowledgements

I want to start by thanking my principal supervisor Patrick Miller for your excellent guidance during the past four years. I thank also Len Thomas, my secondary supervisor for the last two years, for guiding me through the dense forest that is statistics. I have greatly enjoyed witnessing both your brain power in action, and look forward to our future collaborations. I thank Patrick in particular for his trust in my abilities and for encouraging me since the days of being a master's student. I also sincerely thank my two examiners, Jakob Tougaard and Simon Northridge, for your thorough review of this thesis and useful comments.

It has been a great privilege to study in St Andrews, and especially to work with a great group of people that is the 'Miller Lab'. (We never managed to find a suitable name because of the diverse kinds of research going on there). Past and present group members Olga Filatova, Ivan Fidutin, Miguel Neves, Sara Tavares, Kagari Aoki, Tomoko Narazaki, Marjoliene Roos, Saana Isojunno, Nicoletta Biassoni, Hannah Wood made my stay a wonderful experience. Sea Mammal Research Unit (SMRU) is a great collection of people, and I should thank in particular also Simon Moss, Matt Bivens, Paddy Pomeroy, Thomas Götz, Silvana Neves, Cormac Booth, Lucia Martín, Rene Swift, Mark Johnson and Clint Blight for friendships and a range of useful interactions, and past and present members of 'Phil's Lab' next door for lunch time fun; Sanna Kuningas, Marina Costa, Gui Bortolotto, Nadia Ramirez, Christel Reyes, Monica Civil, Lindsay Wilson, Diane Claridge and Charlotte Dunn. I have also been very privileged to have been involved in a number of exiting research projects such as Icelandic-Orcas (www.icelandic-orcas.com) in the beautiful Vestmannaeyjar archipelago and in Grundarfjördur with my better half, the Cetacean social behaviour project in the Azores with Fleur Visser and Machiel Oudejans, and the Mingan Island cetacean study with Christian Ramp.

This PhD research was part of a large collaborative study, the 3S (Sea mammals, Sonar, and Safety) behavioural response study, and I'd especially like to acknowledge principal investigators Petter Kvadsheim, Patrick Miller, Frans-Peter Lam and Peter Tyack who

all bring a special skillset to the table. I thank the members of the exclusive P&P's including Saana Isojunno, Lise Sivle and Charlotte Curé for your friendships and professional advice. The people from the Acoustic and Sonar group at TNO, the Netherlands have made it difficult to break the bonds with my home country completely, and I have always come back with great pleasure to visit friends and colleagues like Sander von Benda-Beckmann, Michael Ainslie, Christ de Jong, Sander van IJsselmuide, Mark van Spellen, Frank Benders and Benoit Quesson. I'm also very grateful to all the other 3S team members, including Rune Hansen, Thomas Sivertsen, Lars Kleivane, René Dekeling, Paul Ensor, Eirik Grønningssæter, Eva Hartvig, Leigh Hickmott, and Machiel Oudejans, and the crews of the research vessel H.U. Sverdrup II who made life on board extremely pleasant.

Many of the ideas developed through interactions with the folks at the Centre for Research into Ecological and Environmental Modelling, in particular with the Multi-study Ocean acoustics Human effects Analysis (MOCHA) project funded by the US Office of Naval Research. Catriona Harris, Stacy DeRuiter, Dina Sadykova, Christophe Laplanche, Tiago Marques, Claudia Faustino, Monique McKenzie, Danielle Harris always made me feel at home whenever I was over at CREEM, even though our second home Iceland kept me from visiting more often. I should also thank Gísli Vikingsson and Páll Jóhnson for showing me the hospitality of the great nation of Iceland.

I was very lucky to have Ron Kastelein of SEAMARCO as a mentor in the earlier days of my PhD, who thought me what working with and caring for animals in the captive setting is all about. I should thank also his wife Brigitte Slingerland, and past and present team members Lean Hoek, Leonie Huijser, Tess van der Drift, Krista Krijger, Wim Verboom, Nancy Jennings, Amy McCleod, Martijn Rambags, Robin Gransier and Arie Smink for your help with data collection and the terrific company during my many visits.

Funding for this PhD was provided by a studentship with matched funding from the UK Natural Environment Research Council and The Netherlands Ministry of Defence (administered by TNO), and I would like to thank particularly René Dekeling for his

support. The $3S^2$ project was funded by the US Office of Naval Research, the Norwegian Ministry of Defence, and The Netherlands Ministry of Defence. The harbour porpoise work was supported by The Netherlands Ministry of Infrastructure and the Environment. The experiments with humpback whales in Norwegian waters were carried out under permits issued by the Norwegian Animal Research Authority (Permit No. S-2011/38782), in compliance with ethical use of animals in experimentation. The research protocol was approved by the University of St Andrews Animal Welfare and Ethics Committee and Woods Hole Oceanographic Institution's Animal Care and Use Committee. Training and data collection with the harbour porpoise were conducted under authorization of The Netherlands Ministry of Economy, Agriculture and Innovation, Department of Nature Management, with Endangered Species Permit FF/75A/2005/048. The research protocol for this laboratory work was approved by the University of St Andrews' School of Biology Ethics Committee.

Finally, I owe a great deal of gratitude to my friends from the Netherlands, my second family from Portugal and my lovely mom and dad that have stood by me throughout this somewhat self-centred journey. The greatest thank you goes to my partner Filipa, for her patience, companionship and love; I could not have done this without you. I dedicate this work to my late sister Judith, who will forever be missed.

Table of Contents

List of Acronyms

3D	three dimensional
AIC	Akaike's information criterion
AIS	Automatic identification system
ANOVA	analysis of variance
BRS	behavioural response study
BUGS	Bayesian inference Using Gibbs Sampling
CEE	controlled exposure experiment
CI	confidence interval or Bayesian credible interval
CR	critical ratio
DAQ	data acquisition
DI	directivity index
DT	detection threshold
EGNOS	European Geostationary Navigation Overlay Service
FFI	Norwegian Defence Research Establishment
FM	frequency modulation
GEE	general estimating equation
GPS	Global Positioning System
HF	high frequency
ID	identifier
IR	infrared
JAGS	Just Another Gibbs Sampler
LED	light-emitting diode
LF	low frequency
LRF	laser range finder
MCMC	Markov chain Monte Carlo
MF	mid-frequency
NL	noise level
PAM	passive acoustic monitoring
PL	propagation loss
QIC	quasi-likelihood under the independence model criterion

R_{min}	minimum source-whale range
RMS	root mean square
RMSE	root mean square error
RT	reaction time
S/N	signal-to-noise ratio
SD	standard deviation
SE	standard error
SEL	sound exposure level
SEL_{cum}	cumulative sound exposure level
SL	source level
SMRU	Sea Mammal Research Unit
SnL	sensation level
SSM	state space model
SPL	sound pressure level
SPL_{max}	maximum sound pressure level (200-ms average)
TNO	Netherlands Organisation for Applied Scientific Research
TTS	temporary threshold shift
UTC	Coordinated Universal Time
UTM	Universal Transverse Mercator
VHF	very high frequency

Chapter 1

General introduction

EFFECTS OF NAVAL SONAR ON CETACEANS

Management of the environmental impacts of anthropogenic activities is an important conservation issue, as the footprint of human influence continues to expand (Beale, 2007; Shannon et al., 2015). Humans have a global impact on geology and ecology in the current epoch, and our activities may threaten the existence of individuals and populations of animals and their ecosystems (Corlett, 2015). Among the principal stressors that can influence marine mammals are climate change, ship strikes, chemical and noise pollution, whaling, and entanglement/entrapment in fishing gear (Thomas et al., 2016). The importance of sound to marine mammals and fishes has led to increased scientific attention for the effects of anthropogenic noise on these animals (Popper and Hawkins, 2016).

Marine mammals generally have acute underwater hearing over a wide frequency range (Mooney et al., 2012; Richardson et al., 1995). They are likely to be vulnerable to noise exposure as they use sound for sexual advertisement (Payne and McVay, 1971), communication (Clausen et al., 2010; Tyack, 1981), feeding (Cerchio and Dahlheim, 2001), orientation (Verfuss et al., 2005) and predator avoidance (Curé et al., 2015). Sources of underwater sound that can have deleterious effects on marine mammals include active sonar during naval exercises (Claridge, 2013; Tyack et al., 2011), propeller noise from vessel traffic (Clark et al., 2009; Rolland et al., 2012), underwater explosions during ship-shock trials, offshore construction and ammunition clearance (Hildebrand, 2009; Ketten et al., 1993; von Benda-Beckmann et al., 2015), pile driving during the construction of offshore wind farms (Brandt et al., 2011; Tougaard et al., 2009) and seismic airguns during surveys for oil and gas exploration (Castellote et al., 2012; McCauley et al., 2000). These sound sources differ in their acoustic characteristics as well as in their spatial and temporal distributions, and may act

independently, cumulatively, or synergistically to negatively influence hearing, behaviour or health of individuals and populations (National Research Council, 2005).

The speed of sound in seawater is about five times that in air, and the absorption of sound energy per unit distance in air is much greater (Kinsler et al., 2000). In contrast, the propagation of electro-magnetic waves (e.g. light) in seawater is far less efficient. As a result, naval fleets use sound waves to detect, locate, and classify submarines, navigate torpedoes and find objects such as mines and other ordnance (Richardson et al., 1995). This suite of techniques is collectively known as sonar (sound navigation and ranging). Passive sonar systems only 'listen' and do not generate sound. Active sonar systems transmit sound pulses under water and time their return echoes to generate 'snapshots' of the environment. Most of these naval active sonars produce high source level pulses, and the high-intensity systems that are used for long- and medium-range detection operate at frequencies below 10 kHz where sound absorption is limited (Ainslie, 2010; Hildebrand, 2009). Therefore, such low- and mid-frequency active sonars may ensonify large volumes of water with sound pressure levels sufficient to be heard by, and possibly disturb, marine life.

National and international legislation reflect the importance of national security and the need for navies to use active sonar. Sonar operators must train under a range of realistic environmental and geographical conditions, which includes training in areas with complex bathymetries. Naval exercises therefore take place in both deep-ocean and coastal waters, and may occur within the natural habitat of many marine mammal species. Concerns about the impacts of active sonar on marine mammals has led to a series of legal disputes instigated by conservation non-governmental organisations which focussed on the US Navy's mitigation measures and environmental impact assessments (McCarty, 2010; Reynolds et al., 2009). These legal cases emphasised the importance of balancing the need for effective sonar training against the risk of harm that active sonar may impose on marine mammals (Zirbel et al., 2011). In a recent settlement in 2015, the US Navy agreed to exclude sonar testing and use of explosives from several biologically important areas in waters off the coast of California and Hawaii.

There was little scientific evidence on the potential impacts of naval active sonar on marine mammals before the late 1990s (Richardson et al., 1995; Simmonds and Lopez-Jurado, 1991). Concerns increased after reports of unusual cetacean mass strandings that coincided in time and space with multi-ship sonar exercises (Balcomb and Claridge, 2001; D'Amico et al., 2009; Fernández et al., 2005; Frantzis, 1998). These mass strandings involved predominantly Cuvier's beaked whales (*Ziphius cavirostris*) and Blainville's beaked whales (*Mesoplodon densirostris*), but the strandings of other cetaceans have also been linked to naval activity (e.g. Balcomb and Claridge, 2001; Hohn et al., 2006; Southall et al., 2006; Yang et al., 2008). The exact chain of events leading to such strandings remains uncertain to date; however, the most plausible mechanisms involve acoustically induced behavioural or physiological responses that lead to stranding directly, or via tissue damage caused by bubble formation (Cox et al., 2006; Hooker et al., 2012). In the Canary Islands, one of the hotspots for beaked whale strandings, no mass standings have occurred since the Spanish government imposed a ban on naval exercises in these waters in 2004 (Férnandez et al., 2013). The use of naval sonar may also have lethal consequences for beaked whales at sea (Fernández et al., 2012) or increase the rates of bycaught harbour porpoises (*Phocoena phocoena*) found stranded (Wright et al., 2013).

There is also anecdotal and experimental evidence that naval sonar can impact behaviour and physiology of cetaceans in a harmful way, without leading to strandings. For example, disruption of normal behaviour of killer whales was observed by whale-watchers and researchers during naval sonar transmissions in Haro Strait, US (NMFS 2005). There were complaints from whale-watchers and fishermen that a multi-ship naval exercise in Vestfjord, Norway caused reduced numbers of killer whales and herring (WWF-Norway 2001). Whilst impacts on herring were later shown to be unlikely (Doksæter et al., 2009; Doksæter et al., 2012; Sivle et al., 2015a), a retrospective analysis of a comparable naval exercise in 2006 suggested that killer whales were displaced from the area in response to a combination of reduced prey availability and sonar use (Kuningas et al., 2013).

Reductions of vocalisations and avoidance by minke whales (*Balaenoptera acutorostrata*) during experimental (Kvadsheim et al., 2015) and actual (Martin et al., 2015) naval sonar exposures were consistent with earlier opportunistic observations of displacement and avoidance behaviour in minke whales during actual sonar exercises (Dolman et al., 2011; Parsons et al., 2000). Vocal responses of humpback whales, *Megaptera novaeangliae* (Fristrup et al., 2003; Miller et al., 2000; Risch et al., 2012) and blue whales, *Balaenoptera musculus* (Melcón et al., 2012) to sonar sounds indicated that such behavioural changes may occur over substantial areas. A recent passive acoustic monitoring (PAM) study conducted during a naval exercise in Danish waters suggested relatively large-scale (~20 km) displacement of harbour porpoises from the area (Tougaard et al., 2015a). The long-term implications of these behavioural changes of marine mammals are currently not well understood; they likely depend on species-specific characteristics of responsiveness, how often and for how long animals are exposed, pressures from other sources of disturbance, and species-specific mechanisms for coping with disturbance.

The lack of understanding about the mechanisms linking sonar to whale strandings and long-term effects of sonar on marine mammals motivated the US Navy and several other navies to fund research on the relationship between sonar exposure and behavioural responses. This ongoing research effort includes experimental studies on laboratory animals, experimental studies on free-ranging animals tagged with multi-sensor digital recording tags (DTAGs; Johnson and Tyack, 2003; Johnson et al., 2009), and (non-experimental) observational studies using long-term satellite tags, visual focal follows or PAM. The primary goals of these studies, collectively known as behavioural response studies (BRSs), is to describe the behavioural responses themselves, to quantify relationships with their potential drivers such as dose and contextual variables, and to study their biological relevance (Harris and Thomas, 2015).

The experimental BRSs predominantly used controlled exposure experiment (CEE) methodology (Tyack et al., 2003), in which controlled doses of the sonar stimulus are applied to test whether or not they induce behavioural (or physiological) responses, which are then contrasted to responses during exposures to control stimuli. CEEs are

aimed at studying fine-scale responses and have provided a wealth of information on various aspects of behavioural responses to naval sonar. Studies on beaked whales have reported strong behavioural responses that were initiated at low received sound levels, with responses consisting of a cessation of foraging, increased swim speed, avoidance and unusual diving patterns (Blainville's beaked whale; Tyack et al., 2011; Cuvier's beaked whale; DeRuiter et al., 2013; Baird's beaked whale, *Berardius bairdii*; Stimpert et al., 2014; northern bottlenose whale, *Hyperoodon ampullatus*; Miller et al., 2015). Despite minor differences in responsiveness across the species, the behavioural responses observed during experimental exposures were in general agreement with patterns of larger scale displacement and changes in echolocation behaviour observed using PAM and satellite tags during real sonar exercises (McCarthy et al., 2011; Moretti et al., 2014; Tyack et al., 2011).

Understanding the effects on *Ziphiidae* will no doubt remain a high research priority given the difficulties in studying these animals and their apparent sensitivity to sonar, although effects on other taxonomic families should not be overlooked. Experimental field studies with other cetacean species have provided evidence that naval sonar can induce behavioural responses of different types, durations, and severities in these species (e.g. Miller et al., 2012; Sivle et al., 2015b). Responses that have been quantified include changes in vocal behaviour (Alves et al., 2014; DeRuiter et al., 2012), changes in activity and energy budgets (Isojunno et al., 2016), changes in social behaviour (Visser et al., 2016), synchronised surfacings with arrivals of sonar pulses (Wensveen et al., 2015a) and behavioural state-specific feeding and displacement responses (Goldbogen et al., 2013b). These field CEEs have also resulted in dose-response functions that reflected species differences and large within- and between animal variation in onset thresholds and response severities (Antunes et al., 2014; Harris et al., 2015b; Miller et al., 2014). Large variation in responsiveness to naval sonar was also observed in small odontocetes on a navy range using opportunistic focal follows (Henderson et al., 2014).

Dose-response functions have also been created from observational (Moretti et al., 2014) and captive research efforts. The captive studies have particularly highlighted the

importance of age, learning/exposure history, species temperament, background noise levels and spectral properties of the signal on animal responsiveness (Houser et al., 2013a; Houser et al., 2013b; Kastelein et al., 2011a; Kastelein et al., 2012). The degree to which behavioural responses in marine mammals match between the captive and free-ranging contexts is uncertain for most species. Observational studies are conducted in the most relevant context, i.e. the actual context in which the risk of harm may occur, but they have no control over the sound stimuli or study animals and generally use lower resolution observational methods that are more likely to miss important changes in behaviour. Captive and observational studies represent opposite trade-offs between realism and control; experimental studies with free-ranging animals can be seen as the intermediate between these two types of study. Which of these three approaches will be the most appropriate in a given situation depends on the research question. They are likely to complement each other, as the strength of one approach often corrects for the weakness of the other (Tyack et al., 2003).

The fates of populations rather than individuals are eventually the most important for achieving sustainable conservation outcomes (Hatch and Fristrup, 2009). Behavioural and sublethal physiological disturbance from sonar exposure may affect an individual's health, survival, or ability to reproduce, so these effects may affect population growth rates and ecosystem dynamics (National Research Council, 2005). As such, sonar exposure may become biologically significant if it induces changes in energy intake or expenditure (Pirotta et al., 2015), social disruption, or avoidance of an area that is important for life functions such as feeding, breeding, or resting (New et al., 2013a). Although direct and immediate effects of naval sonar (e.g. hearing loss, the separation of a dependent calf from its mother) may impact individual health, there is currently little knowledge about the temporal and spatial extent over which noise disturbance needs to occur to result in significant population consequences. Long-term behavioural effects are difficult to observe in free-ranging marine mammals, and these responses could be modified by learning (Bejder et al., 2006, 2009), acoustic masking (e.g. Clark et al., 2009), hearing loss (Finneran, 2015) and stress responses (Atkinson et al., 2015).

A model framework has been developed for predicting population-level effects in marine mammals from short-term behavioural or physiological responses, known as the Population Consequences of Disturbance (PCoD) framework (New et al., 2013b), after an earlier conceptual version was presented (Lusseau and Bejder, 2007; National Research Council, 2005). This modelling approach has been successfully applied in a number of specific case studies for which enough empirical data were available to parameterise the model (New et al., 2013b; New et al., 2014; Pirotta et al., 2015; Schick et al., 2013) and has also been applied using information obtained from expert-elicitation (King et al., 2015). Sonar BRSs generally measure relatively short-term behavioural responses, of which some are expected to have biologically significant effects if exposures would continue for longer than actually realised during the experiments. However, the detailed information on behavioural responses (e.g. changes in energy balance) collected by these experimental studies might be used in the future to partly parameterise PCoD-type models, in order to predict risks of population-level effects.

MANAGEMENT OF ENVIRONMENTAL RISK OF SONAR

In many countries, legislation aimed at protecting marine life requires the assessment of human activities that may produce harmful levels of underwater noise (Roman et al., 2013; Tasker et al., 2010). It is also often required to prepare environmental impact assessments or statements for proposed activities that generate noise. If the assessment finds that the activity poses a significant environmental risk, use of risk mitigation methods and/or restrictions on operations may be required. In contrast to other noise producers, however, navies typically self-regulate their potential environmental impacts and set their own mitigation protocols (Dolman et al., 2009; Dolman et al., 2011).

Both qualitative and quantitative methods for assessing the environmental risks of anthropogenic noise generally follow the conceptual framework developed by the US Environmental Protection Agency (EPA, 1992) for impacts such as chemical pollution (for details on applying this framework to underwater noise, see Boyd et al., 2008; Harwood, 2000). In the context of marine mammals, such quantitative methods generally use individual-based modelling techniques to construct the exposure histories

of simulated animals that move through virtual sound fields and then evaluate how often received levels reach certain risk thresholds (Frankel et al., 2002). A number of navies currently use this approach in the planning stage of naval activities to avoid important marine mammal habitats and thereby minimise risks (e.g. Donovan et al., 2012; Nordlund and Kvadsheim, 2014). Environmental Impact Statements of the US Navy have also used comparable methods to estimate the number of animals that are affected behaviourally or physiologically by naval sonar and other sounds (Schecklman et al., 2011; Wartzok et al., 2012) for compliance with US environmental legislation. The risk thresholds that are used in these various assessments greatly affect the estimated impact of the noise.

Sound exposure guidelines for detrimental effects such as hearing injury (i.e. permanent hearing threshold shift; PTS), temporary threshold shift (TTS) and behavioural disturbance have been proposed for marine mammals (Southall et al., 2007) as well as fishes and sea turtles (Popper et al., 2014). These guidelines have considered groups of sound sources based on their acoustic properties and groups of species based on their expected hearing abilities. The risk thresholds in these guidelines are preferably expressed as numeric values that represent the received sound levels at which impacts are expected to occur; however, scientific knowledge has not always been sufficient for the determination of such acoustic thresholds. For example, Southall et al. (2007) and the most recent draft guidelines for acoustic exposure by the US National Oceanographic and Atmospheric Administration (NOAA, 2015) were unable to establish noise criteria for behavioural disturbance because of the lack of scientific consensus. Acoustic thresholds can either be represented by a step function (i.e. instant change of risk from 0 to 100 % at one sound level) or a probabilistic dose-response function (i.e. the risk gradually increases with sound level). Because behavioural effects are expected to occur over a relatively wide range of received sound levels, the US Navy and other navies have used dose-response functions for predicting behavioural effects in marine mammals. These functions are currently being updated based on recent experimental information collected by BRSs (Henderson, 2015).

The acoustic risk thresholds in sound exposure criteria are generally expressed as sound levels that are ‘frequency-weighted’ according to the perception of the animals.

Frequency weighting greatly simplifies noise criteria because it results in single thresholds that apply to many sounds irrespective of their frequency spectra. Weighting functions represent the average frequency response of a species' auditory system; they attempt to emulate how sounds are perceived by the animal. Therefore, weighting functions are useful for the extrapolation of observed risk thresholds (e.g. for the onset of TTS or avoidance responses) to frequencies outside of the range for which the responses were measured.

Different types of weighting functions are currently available for marine mammals: 1) weighting functions based upon the audiogram (e.g., Miller et al., 2014; Verboom and Kastelein, 2005), 2) weighting functions that are flat within the effective frequency range of hearing, known as M-weighting (Southall et al., 2007), 3) weighting functions derived from equal loudness and equal latency contours (Finneran and Schlundt, 2011; Finneran and Schlundt, 2013; Kastelein et al., 2011b), and 4) weighting functions that are based upon a combination of these types of information (NOAA, 2015). There are significant differences in shape between these functions, so they may lead to very different estimates of risk of harm (e.g. de Jong and Ainslie, 2008).

Weighting functions used in the assessment of effects of noise on humans are based on equal loudness contours (e.g. A- and C-weighting functions, Kinsler et al., 2000), and therefore functions describing loudness perception in marine mammals may also be effective for estimating effects of sounds of mid- to high-intensity in marine mammals. However, perceived loudness is a subjective descriptor of sound level, which makes it difficult to quantify in animals. Finneran and Schlundt (2011) measured the equal loudness contours of a single bottlenose dolphin (*Tursiops truncatus*) following a loudness comparison method (Suzuki and Takeshima, 2004). It was very difficult to convey the complex task to the dolphin, and more than 15,000 individual trials had to be completed. However, equal latency contours, which describe the frequency-dependent relationships between received sound level and reaction time to a sound stimulus, are roughly similar to equal loudness contours in humans (Marshall and Brandt, 1980; Pfingst et al., 1975a; Pfingst et al., 1975b). Such contours may therefore be useful as a proxy because reaction time is much easier to measure for animals than loudness. Equal

latency contours have been obtained for the macaque (Stebbins, 1966), squirrel monkey (Green, 1975), house finch (Dooling et al., 1978) and domestic cat (May et al., 2009); preliminary data exist for a few marine mammal species (Kastelein et al., 2011b; Mulsow and Finneran, 2013; Reichmuth, 2013; Ridgway and Carder, 2000). Frequency weighting based on equal latency contours may therefore provide a good alternative when equal loudness data are not available.

Best-practise guidelines for human activities that generate noise in the marine environment currently recommend the use of mitigation methods designed to minimise risk of harm to marine mammals (e.g. Compton et al., 2008). Mitigation methods for naval sonar can be broadly categorised as 1) time/area planning of exercises to avoid important marine mammal habitat, 2) implementation of operational procedures such as shut-down based upon PAM and visual monitoring of animals for the purpose of maintaining an 'exclusion zone' around the source (Dolman et al., 2009; Dolman et al., 2011). One of the most common operational mitigation procedures is the gradual increase of source intensity prior to normal (full-power) operation, a procedure 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 sound sources such as seismic surveys (Gordon et al., 2003; Nowacek et al., 2013), acoustic thermometry (Frankel and Clark, 2000a), and pile driving and detonation of explosions (Brandt et al., 2011; dos Santos et al., 2010; Jefferson et al., 2009).

The assumption behind the ramp-up mitigation procedure is that animals will move away from the path of an approaching sound source or the location of a stationary source as the source level gradually increases, thus reducing the maximum sound intensity and energy they receive. Ramp-up of naval sonar is intended to mitigate against auditory (Mooney et al., 2009) and other types of physiological damage in marine mammals that are relative 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 power. This mitigation method is generally thought to be a 'common sense' procedure (Stone and Tasker 2006, Weir and Dolman 2007); however, there is currently little scientific evidence whether ramp-up actually works as intended or not. In theory, ramp-up might do more harm than good by attracting animals to the source when levels

are low, or by increasing the length of time the animals are exposed (Barlow and Gisiner, 2006). Some argue that ramp-up affects the fidelity of naval combat training. It is therefore important to experimentally evaluate the effectiveness of ramp-up as a method to mitigate risk to marine mammals during sonar operations.

There are some reports of cetaceans initiating avoidance responses during ramp-up, although these reports do not quantify the effect of the avoidance response on the risk of harm. Avoidance sometimes occurred during the ramp-up period of controlled experiments with an approaching source vessel in which free-ranging killer whales, *Orcinus orca*, long-finned pilot whales, *Globicephala melas*, and sperm whales, *Physeter microcephalus* were exposed to 1-2 and 6-7 kHz naval sonar (Miller et al., 2012). Weir (2008a) observed a pod of Atlantic spotted dolphins approaching with the apparent intent to bow-ride during a seismic survey in Gabon. Early in the ramp-up the animals suddenly veered away from the vessel while at ~500 m from the source. During seismic ramp-up in Angola, Weir (2008b) observed a group of short-finned pilot whales making a sharp turn away from the path of the vessel, but the avoidance was more limited in time and space. A non-peer-reviewed correlational study based on observational data collected over 7 years also provided evidence that seismic ramp-up indeed triggers avoidance behaviour in cetaceans (Stone, 2015). Von Benda-Beckmann and colleagues (2014) conducted a simulation study and found that ramp-up of naval sonar can be effective at reducing the number of animals experiencing sound energy that is high enough to cause temporary or permanent hearing loss. Important factors were the assumed dose-response relationship and swimming speed of the animals, as well as the ramp-up duration, sailing speed and time interval between the sonar pulses (von Benda-Beckmann et al., 2014, 2016). Even though the simulation study used relevant empirical input in its model, experimental confirmation of these predictions has still been missing.

THESIS OUTLINE

In this thesis, I studied the potential effects of naval sonar on cetaceans by means of novel experimental and computational methods, in order to inform criteria for allowable levels of naval sonar. In Chapters 2, 3 and 5, I used multi-variate data sets for North-

Atlantic humpback whales which were collected as part of the $3S^2$ BRS project in high-latitude waters [Figure 1.1; Kvadsheim et al. (2015); Lam et al. (2016)]. In Chapter 4, the remaining data chapter, I conducted a psychoacoustic study with a trained harbour porpoise at the SEAMARCO Research Institute (Kastelein et al., 2009). The main findings of this thesis are integrated in Chapter 6, where also the broader management implications of the work are discussed. A short description of the goals and methods of each data chapter is given below; results and conclusions are summarised at the start of each of these chapters. This thesis addressed several of the recommendations made in a recent review (Shannon et al., 2015); that is, to measure responses over a gradient of noise levels, to evaluate mitigation measures, and to improve reporting of acoustic metrics.

Chapter 2. A method for reconstructing fine-scale tracks from dead-reckoning and position fixes applied to humpback whales

The goal of chapter 2 was to improve the accuracy of movement tracks of humpback whales tagged with multi-sensor data loggers. To achieve this, a novel state-space method was developed that integrates information on horizontal positions (Fastloc-GPS), body orientation and flow noise (DTAG), and range and bearing estimates made by visual observers during focal follows. Models were fitted in the Bayesian statistical framework to account for differences in the distributions of the data, time resolutions and periods without observations. The method was applied to all humpback whales in the data set, and within- and between-animal variability in the movement tracks was described.

Aims and objectives

- To develop a state-space model framework for reconstructing whale tracks from dead-reckoning, Fastloc-GPS, and visual observations.
- To quantify the spatial accuracy of the Fastloc-GPS and visual (range and bearing) observations in dedicated tests.
- To reconstruct the movement tracks of all humpback whales in the $3S^2$ data set.

Chapter 3. Dose-response functions for avoidance onset and feeding cessation by humpback whales in response to naval sonar

In this chapter, several dose-response relationships were generated to relate different aspects of naval sonar exposure (i.e. amplitude, energy, distance) to the probability of onset of two behavioural response types; horizontal avoidance and cessation of feeding. The data analysed in this chapter were collected during experimental vessel approaches on tagged humpback whales, which were conducted either with or without active transmissions of 1.3-2 kHz naval sonar. The source positions, whale positions (i.e. the tracks created in Chapter 2) and received sound levels during these experiments were evaluated to determine the response onset thresholds of the humpback whales. Additionally, the outcomes were compared to other marine mammal dose-response functions published to date and differences with the recently proposed method of Harris et al. (2015) were investigated.

Aims and objectives

- To measure the response onset thresholds of the humpback whales.
- To construct dose-response functions for avoidance onset and feeding cessation.
- To formally investigate the potential effects of behavioural state and exposure history on humpback whale responsiveness.
- To informally assess the potential effects of species, methodological differences, and other factors on behavioural responsiveness.

Chapter 4. Auditory weighting functions based on behavioural reaction times of a harbour porpoise

In chapter 4, the behavioural reaction times of a harbour porpoise to tonal sounds were measured in a laboratory study using operant conditioning and a psychoacoustic method. Contours of equal reaction time across signal frequency and received sound pressure level were converted into auditory weighting functions that can be used as a proxy for the frequency response of the animal's auditory system. The weighting functions may therefore be used to more accurately predict loudness perception in harbour porpoises. For validation, the derived weighting functions were compared to published onset thresholds for behavioural and hearing effects in harbour porpoises.

Aims and objectives

- To measure the reaction times of a harbour porpoise to sound signals with a wide range of frequencies and amplitudes.
- To construct auditory weighting functions from equal latency contours.
- To compare the weighting functions to published onset thresholds for behavioural and hearing effects in harbour porpoises.

Chapter 5. Experimental evaluation of the effectiveness of ramp-up as a mitigation method for naval sonar

In this chapter, the humpback whale was used as a model species to evaluate an operational method for mitigating acute effects from sonar in marine mammals. The same experiments with humpback whales analysed in Chapter 3 were used to address this goal. These experiments were specifically designed to test the effectiveness of ramp-up by comparing received levels during full-power sonar operations with and without transmission of a preceding ramp-up scheme. The overarching goal was to increase understanding about the biological factors that influence ramp-up effectiveness in marine mammals in general. I used a regression analysis based on generalised estimating equations to model the effects of the treatment on the received sound levels and minimum source-whale distance during experimental sessions.

Aims and objectives

- To quantify the potential effect of ramp-up on the sound levels received by humpback whales relative to no ramp-up
- To investigate the biological factors affecting the effectiveness of ramp-up

REFERENCES

Ainslie, M. A. (2010). Principles of Sonar Performance Modeling. Chichester, UK: Springer-Praxis.

Alves, A., Antunes, R., Bird, A., Tyack, P. L., Miller, P. J. O., Lam, F.-P. A. and Kvadsheim, P. H. (2014). Vocal matching of naval sonar signals by long-finned pilot whales (*Globicephala melas*). Mar. Mammal Sci. 30, 1248–1257.

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

Atkinson, S., Crocker, D., Houser, D. and Mashburn, K. (2015). Stress physiology in marine mammals: how well do they fit the terrestrial model? J. Comp. Physiol. B 1–24.

Balcomb, K. C. and Claridge, D. E. (2001). A mass stranding of cetaceans caused by naval sonar in the Bahamas. Bahamas J. Sci. 5, 2–12.

Barlow, J. and Gisiner, R. (2006). Mitigating, monitoring and assessing the effects of anthropogenic sound on beaked whales. J. Cetacean Res. Manag. 7, 239–249.

Beale, C. M. (2007). The behavioral ecology of disturbance responses. Int. J. Comp. Psychol. 20, 111–120.

Bejder, L., Samuels, A., Whitehead, H. and Gales, N. (2006). Interpreting short-term behavioural responses to disturbance within a longitudinal perspective. Anim. Behav. 72, 1149–1158.

Bejder, L., Samuels, A., Whitehead, H., Finn, H. and Allen, S. (2009). Impact assessment research: use and misuse of habituation, sensitisation and tolerance in describing wildlife responses to anthropogenic stimuli. Mar. Ecol. Prog. Ser. 395, 177–185.

Boyd, I., Brownell, B., Cato, D., Clark, C., Costa, D., Evans, P., Gedamke, J., Gentry, R., Gisiner, R., Gordon, J., et al. (2008). The effects of anthropogenic sound on marine mammals. A draft research strategy. (ed. Connolly, N.) Marine Board-European Science Foundation.

Brandt, M. J., Diederichs, A., Betke, K. and Nehls, G. (2011). Responses of harbour porpoises to pile driving at the Horns Rev II offshore wind farm in the Danish North Sea. Mar. Ecol. Prog. Ser. 421, 205–216.

British Oceanographic Data Centre (2010). The GEBCO_08 Grid.

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

Cerchio, S. and Dahlheim, M. (2001). Variation in feeding vocalizations of humpback whales *Megaptera novaeangliae* from southeast Alaska. Bioacoustics 11, 277–295.

Claridge, D. (2013). Population Ecology of Blainville's Beaked Whales (*Mesoplodon densirostris*). PhD Thesis, University of St Andrews, St Andrews, UK.

Clark, C., Ellison, W. T., Southall, B. L., Hatch, L., Van Parijs, S. M., Frankel, A. and Ponirakis, D. (2009). Acoustic masking in marine ecosystems: intuitions, analysis, and implication. Mar. Ecol. Prog. Ser. 395, 201–222.

Clausen, K. T., Wahlberg, M., Beedholm, K., Deruiter, S. and Madsen, P. T. (2010). Click Communication in Harbour Porpoises *Phocoena Phocoena*. Bioacoustics 20, 1–28.

Compton, R., Goodwin, L., Handy, R. and Abbott, V. (2008). A critical examination of worldwide guidelines for minimising the disturbance to marine mammals during seismic surveys. Mar. Pol. 32, 255–262.

Corlett, R. T. (2015). The Anthropocene concept in ecology and conservation. Trends Ecol. Evol. 30, 36–41.

Cox, T. M., Ragen, T. J., Read, A. J., Vos, E., Baird, R. W., Balcomb, K., Barlow, J., Caldwell, J., Cranford, T., Crum, L., et al. (2006). Understanding the impacts of anthropogenic sound on beaked whales. J. Cetacean Res. Manag. 7, 177–187.

Curé, C., Doksaeter 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.

D’Amico, A., Gisiner, R. C., Ketten, D. R., Hammock, J. A., Johnson, C., Tyack, P. L. and Mead, J. (2009). Beaked whale strandings and naval exercises. Aquat. Mamm. 35, 452–472.

de Jong, C. A. F. and Ainslie, M. A. (2008). Underwater radiated noise due to the piling for the Q7 Offshore Wind Park. In Proc. of Acoustics ’08, Paris, France.

DeRuiter, S. L., Boyd, I. L., Claridge, D. E., Clark, C. W., Gagnon, C., Southall, B. L. and Tyack, P. L. (2012). Delphinid whistle production and call matching during playback of simulated military sonar. Mar. Mammal Sci. no–no.

DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone, E. A., Friedlaender, A. S., Joseph, J. E., Moretti, D., Schorr, G. S., et al. (2013). First direct measurements of behavioural responses by Cuvier’s beaked whales to mid-frequency active sonar. Biol. Lett. 9, 20130223.

Doksæter, L., Godø, O. R., Olsen, E., Nøttestad, L. and Patel, R. (2009). Ecological studies of marine mammals using a seabed-mounted echosounder. ICES J. Mar. Sci. 66, 1029–1036.

Doksæter, L., Handegard, N. O., Godø, O. R., Kvadsheim, P. H. and Nordlund, N. (2012). Behavior of captive herring exposed to naval sonar transmissions (1.0-1.6 kHz) throughout a yearly cycle. J. Acoust. Soc. Am. 131, 1632–42.

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.

Dolman, S. J., Evans, P. G., Notarbartolo-di-Sciara, G. and Frisch, H. (2011). Active sonar, beaked whales and European regional policy. Mar. Pollut. Bull. 63, 27–34.

Donovan, C. R., Harris, C., Harwood, J. and Milazzo, L. (2012). A simulation-based method for quantifying and mitigating the effects of anthropogenic sound on marine mammals. In Proceedings of Meetings on Acoustics, p. 070043.

Dooling, R. J., Zoloth, S. R. and Baylis, J. R. (1978). Auditory sensitivity, equal loudness, temporal resolving power, and vocalizations in the house finch (*Carpodacus mexicanus*). J. Comp. Physiol. Psychol. 92, 867–876.

dos Santos, M. E., Couchinho, M. N., Rita Luis, A. and Goncalves, E. J. (2010). Monitoring underwater explosions in the habitat of resident bottlenose dolphins. J. Acoust. Soc. Am. 128, 3805–3808.

EPA (1992). Framework for Ecological Risk Assessment. Washington, DC.

Fernández, A., Edwards, J. F., Rodríguez, F., Espinosa de los Monteros, A., Herráez, P., Castro, P., Jaber, J. R., Martín, V. and Arbelo, M. (2005). "Gas and fat embolic syndrome" involving a mass stranding of beaked whales (Family *Ziphiidae*) exposed to anthropogenic sonar signals. Vet. Pathol. 42, 446–457.

Fernández, A., Sierra, E., Artin, V., Méndez, M., Sacchinni, S., Bernaldo de Quirós, Y., Andrada, M., Rivero, M., Quesada, O., Tejedor, M., et al. (2012). Last "atypical" beaked whales mass stranding in the Canary Islands (July, 2004). J. Mar. Sci. Res. Dev. 02, 1000107.

Férnandez, A., Arbelo, M. and Martin, V. (2013). No mass strandings since sonar ban. Nature 497, 317.

Finneran, J. J. (2015). Noise-induced hearing loss in marine mammals: A review of temporary threshold shift studies from 1996 to 2015. J. Acoust. Soc. Am. 138, 1702–1726.

Finneran, J. J. and Schlundt, C. E. (2011). Subjective loudness level measurements and equal loudness contours in a bottlenose dolphin (*Tursiops truncatus*). J. Acoust. Soc. Am. 130, 3124–3136.

Finneran, J. J. and Schlundt, C. E. (2013). Effects of fatiguing tone frequency on temporary threshold shift in bottlenose dolphins (*Tursiops truncatus*). J. Acoust. Soc. Am. 133, 1819–1826.

Frankel, A. S. and Clark, C. W. (2000). Behavioral responses of humpback whales (*Megaptera novaeangliae*) to full-scale ATOC signals. J. Acoust. Soc. Am. 108, 1930–1937.

Frankel, A. S., Ellison, W. T. and Buchanan, J. (2002). Application of the Acoustic Integration Model (AIM) to predict and minimize environmental impacts. In Oceans '02 MTS/IEEE, pp. 1438–1443.

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

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

Goldbogen, J. A., 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.

Gordon, J., Gillespie, D., Potter, J., Frantzis, A., Simmonds, M. P., Swift, R. and Thompson, D. (2003). A review of the effects of seismic surveys on marine mammals. Mar. Tech. Soc. J. 37, 16–34.

Green, S. (1975). Auditory sensitivity and equal loudness in the squirrel monkey (*Saimiri sciureus*). J. Exp. Anal. Behav. 23, 255–264.

Harris, C. M. and Thomas, L. (2015). Status and future of research on the behavioural responses of marine mammals to U.S. Navy sonar. CREEM Technical Report 2015-3, University of St Andrews, UK.

Harris, C. M., Sadykova, D., DeRuiter, S. L., Tyack, P. L., Miller, P. J. O., Kvadsheim, P. H., Lam, F. P. A. and Thomas, L. (2015). Dose response severity functions for acoustic disturbance in cetaceans using recurrent event survival analysis. Ecosphere 6, art236.

Harwood, J. (2000). Risk assessment and decision analysis in conservation. Biol. Cons. 95, 219–226.

Hatch, L. T. and Fristrup, K. M. (2009). No barrier at the boundaries: implementing regional frameworks for noise management in protected natural areas. Mar. Ecol. Prog. Ser. 395, 223–244.

Henderson, E. (2015). The use of behavioral response study data in the development of behavioral risk functions for the US Navy. In Abstracts of the Sea Mammals and Sonar Symposium, 27-28 October 2015, University of St Andrews.

Henderson, E. E., Smith, M. H., Gassmann, M., Wiggins, S. M., Douglas, A. B. and Hildebrand, J. a (2014). Delphinid behavioral responses to incidental mid-frequency active sonar. J. Acoust. Soc. Am. 136, 2003–14.

Hildebrand, J. A. (2009). Anthropogenic and natural sources of ambient noise in the ocean. Mar. Ecol. Prog. Ser. 395, 5–20.

Hohn, A. A., Rotstein, D. S., Harms, C. A. and Southall, B. L. (2006). Report on marine mammal unusual mortality event UMESE-0501Sp: multispecies mass stranding of pilot whales (*Globicephala macrorhynchus*), minke whale (*Balaenoptera acutorostrata*), and dwarf sperm whales (*Kogia sima*) in North Carolina on 15-16 January 2005. Silver Spring, MD: National Marine Fisheries Service.

Hooker, S. K., Fahlman, a, Moore, M. J., de Soto, N. A., de Quirós, Y. B., Brubakk, a O., Costa, D. P., Costidis, a M., Dennison, S., Falke, K. J., et al. (2012). Deadly diving? Physiological and behavioural management of decompression stress in diving mammals. Proc. Biol. Sci. 279, 1041–50.

Houser, D. S., Martin, S. W. and Finneran, J. J. (2013a). Exposure amplitude and repetition affect bottlenose dolphin behavioral responses to simulated mid-frequency sonar signals. J. Exp. Mar. Bio. Ecol. 443, 123–133.

Houser, D. S., Martin, S. W. and Finneran, J. J. (2013b). Behavioral responses of California sea lions to mid-frequency (3250-3450 Hz) sonar signals. Mar. Environ. Res. 92, 268–278.

Isojunno, S., Curé, C., Kvadsheim, P., Lam, F., Tyack, P., Wensveen, P. and Miller, P. (2016). Sperm whales reduce foraging effort during exposure to 1-2 kHz sonar and killer whale sounds. Ecol. Appl. In press.

Jefferson, T. A., Hung, S. K. and Wursig, B. (2009). Protecting small cetaceans from coastal development: Impact assessment and mitigation experience in Hong Kong. Mar. Pol. 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.

Johnson, M., Aguilar de Soto, N. and Madsen, P. T. (2009). Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: a review. Mar. Ecol. Prog. Ser. 395, 55–73.

Kastelein, R. A., Wensveen, P. J., Hoek, L., Au, W. W., Terhune, J. M. and de Jong, C. A. (2009). Critical ratios in harbor porpoises (*Phocoena phocoena*) for tonal signals between 0.315 and 150 kHz in random Gaussian white noise. J Acoust Soc Am 126, 1588–1597.

Kastelein, R. A., Steen, N., de Jong, C., Wensveen, P. J. and Verboom, W. C. (2011a). Effect of broadband-noise masking on the behavioral response of a harbor

porpoise (*Phocoena phocoena*) to 1-s duration 6-7 kHz sonar up-sweeps. J. Acoust. Soc. Am. 129, 2307–2315.

Kastelein, R. A., Wensveen, P. J., Terhune, J. M. and de Jong, C. A. (2011b). Near-threshold equal-loudness contours for harbor seals (*Phoca vitulina*) derived from reaction times during underwater audiometry: a preliminary study. J. Acoust. Soc. Am. 129, 488–495.

Kastelein, R. A., Steen, N., Gransier, R., Wensveen, P. J. and de Jong, C. A. (2012). Threshold received sound pressure levels of single 1-2 kHz and 6-7 kHz up-sweeps and down-sweeps causing startle responses in a harbor porpoise (*Phocoena phocoena*). J. Acoust. Soc. Am. 131, 2325–2333.

Ketten, D. R., Lien, J. and Todd, S. (1993). Blast injury in humpback whale ears: evidence and implications. J. Acoust. Soc. Am. 94, 1849–1850.

King, S. L., Schick, R. S., Donovan, C., Booth, C. G., Burgman, M., Thomas, L. and Harwood, J. (2015). An interim framework for assessing the population consequences of disturbance. Methods Ecol. Evol. 6, 1150–1158.

Kinsler, L. E., Frey, A. R., Coppens, A. B. and Sanders, J. V (2000). Fundamentals of Acoustics. 4th Editio. Hoboken, NJ: John Wiley & Sons.

Kuningas, S., Kvadsheim, P. H., Lam, F.-P. A. and Miller, P. J. O. (2013). Killer whale presence in relation to naval sonar activity and prey abundance in northern Norway. ICES J. Mar. Sci. 70, 1287–1293.

Kvadsheim, P. H., Lam, F., Miller, P., Sivle, L. D., Wensveen, P., Roos, M., Tyack, P., Kleivane, L., Visser, F., Curé, C., et al. (2015). The 3S2 experiments - studying the behavioral effects of naval sonar on northern bottlenose whales, humpback whales and minke whales. FFI Report 2015/01001.

Lam, F.-P. A., Kvadsheim, P. H., Miller, P. J. O., Tyack, P. L., Ainslie, M. A., Curé, C., Kleivane, L., Sivle, L. D., van Ijsselmuide, S. P., Visser, F., et al. (2016). Controlled Sonar Exposure Experiments on Cetaceans in Norwegian Waters: Overview of the 3S-Project. In The Effects of Noise on Aquatic Life II, Advances in Experimental Medicine and Biology, pp. 589–598.

Lusseau, D. and Bejder, L. (2007). The long-term consequences of short-term responses to disturbance experiences from whalewatching impact assessment. Int. J. Comp. Psychol. 20, 228–236.

Marshall, L. and Brandt, J. F. (1980). The relationship between loudness and reaction time in normal hearing listeners. Acta Otolaryngol. 90, 244–249.

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

May, B. J., Little, N. and Saylor, S. (2009). Loudness perception in the domestic cat: reaction time estimates of equal loudness contours and recruitment effects. J Assoc. Res. Otolaryngol. 10, 295–308.

McCarthy, E., Moretti, D., Thomas, L., DiMarzio, N., Morrissey, R., Jarvis, S., Ward, J., Izzi, A. and Dilley, A. (2011). Changes in spatial and temporal distribution and vocal behavior of Blainville's beaked whales (*Mesoplodon densirostris*) during multiship exercises with mid-frequency sonar. Mar. Mammal Sci. 27, E206–E226.

McCarty, R. T. (2010). Note: Winter v. NRDC: The Navy, submarines, active sonar, and whales - an analysis of the ninth circuit review and the roberts court extension of the military deference doctrine. Hous. L. Rev. 47, 489–527.

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.

Melcón, M. L., Cummins, A. J., Kerosky, S. M., Roche, L. K., Wiggins, S. M. and Hildebrand, J. a (2012). Blue whales respond to anthropogenic noise. PLoS One 7, e32681.

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

Miller, P. J. 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.

Miller, P. J. O., Antunes, R. N., Wensveen, P. J., Samarra, F. I. P., Catarina Alves, A., Tyack, P. L., Kvadsheim, P. H., Kleivane, L., Lam, F.-P. A., Ainslie, M. A., et

al. (2014). Dose-response relationships for the onset of avoidance of sonar by free-ranging killer whales. J. Acoust. Soc. Am. 135, 975–993.

Miller, P. J. O., Kvadsheim, P. H., Lam, F. P. A., Tyack, P. L., Curé, C., DeRuiter, S. L., Kleivane, L., Sivle, L. D., van IJsselmuide, S. P., Visser, F., et al. (2015). First indications that northern bottlenose whales are sensitive to behavioural disturbance from anthropogenic noise. R. Soc. Open Sci. 2, 140484.

Mooney, T. A., Nachtigall, P. E. and Vlachos, S. (2009). Sonar-induced temporary hearing loss in dolphins. Biol. Lett. 5, 565–567.

Mooney, T. A., Yamato, M. and Branstetter, B. K. (2012). Hearing in cetaceans: from natural history to experimental biology. Adv. Mar. Biol. 63, 197–246.

Moretti, D., Thomas, L., Marques, T., Harwood, J., Dilley, A., Neales, B., Shaffer, J., McCarthy, E., New, L., Jarvis, S., 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.

Mulsow, J. and Finneran, J. J. (2013). Auditory reaction time measurements and equal-latency curves in the California sea lion (*Zalophus californianus*) and bottlenose dolphin (*Tursiops truncatus*). Proc. Meet. Acoust. 19, 010003.

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

New, L. F., Moretti, D. J., Hooker, S. K., Costa, D. P. and Simmons, S. E. (2013a). Using energetic models to investigate the survival and reproduction of beaked whales (family *Ziphiidae*). PLoS One 8, e68725.

New, L. F., Harwood, J., Thomas, L., Donovan, C., Clark, J. S., Thompson, P. M., Cheney, B., Scott-Hayward, L. and Lusseau, D. (2013b). Modelling the biological significance of behavioural change in coastal bottlenose dolphins in response to disturbance. Funct. Ecol. 27, 314–322.

New, L., Clark, J., Costa, D., Fleishman, E., Hindell, M., Klanjšček, T., Lusseau, D., Kraus, S., McMahon, C., Robinson, P., et al. (2014). Using short-term measures of behaviour to estimate long-term fitness of southern elephant seals. Mar. Ecol. Prog. Ser. 496, 99–108.

NOAA (2015). DRAFT Guidance for Assessing the Effects of Anthropogenic Sound on

Marine Mammal Hearing - Underwater Acoustic Threshold Levels for Onset of Permanent and Temporary Threshold Shifts. Maryland, MD: US Department of Commerce.

Nordlund, N. and Kvadsheim, P. (2014). SONATE 2015 – a decision aid tool to mitigate the impacts of sonar operations on marine life. Horten, Norway.

Nowacek, D. P., Bröker, K., Donovan, G., Gailey, G., Racca, R., Reeves, R. R., Vedenev, A. I., Weller, D. W. and Southall, B. L. (2013). Responsible practices for minimizing and monitoring environmental impacts of marine seismic surveys with an emphasis on marine mammals. Aquat. Mamm. 39, 356–377.

Parsons, E. C. M., Birks, I., Evans, P. G. H., Gordon, J. C. D., Shrimpton, J. H. and Pooley, S. (2000). The possible impacts of military activity on cetaceans in west Scotland. Eur. Res. Cetaceans 14, 185–190.

Payne, R. S. and McVay, S. (1971). Songs of humpback whales. Humpbacks emit sounds in long, predictable patterns ranging over frequencies audible to humans. Science 173, 585–597.

Pfingst, B. E., Hienz, R., Kimm, J. and Miller, J. (1975a). Reaction-time procedure for measurement of hearing. I. Suprathreshold functions. J. Acoust. Soc. Am. 57, 421–430.

Pfingst, B. E., Hienz, R. and Miller, J. (1975b). Reaction-time procedure for measurement of hearing. II. Threshold functions. J. Acoust. Soc. Am. 57, 431–436.

Pirotta, E., Merchant, N. D., Thompson, P. M., Barton, T. R. and Lusseau, D. (2015). Quantifying the effect of boat disturbance on bottlenose dolphin foraging activity. Biol. Conserv. 181, 82–89.

Popper, A. N. and Hawkins, A. (2016). The Effects of Noise on Aquatic Life II. New York, NY: Springer Science+Business Media.

Popper, A. N., Hawkins, A. D., Fay, R. R., Mann, D. A., Bartol, S., Carlson, T. J., Coombs, S., Ellison, W. T., Gentry, R. L., Halvorsen, M. B., et al. (2014). Sound exposure guidelines for fishes and sea turtles.

Reichmuth, C. J. (2013). Equal loudness contours and possible weighting functions for pinnipeds. J. Acoust. Soc. Am. 134, 4210.

Reynolds, J. R., Kiekow, T. G. and Zak Smith, S. (2009). No whale of a tale: legal implications of Winter v. NRDC. Ecol. Law Q. 36, 753–774.

Richardson, W. J., Greene, C. R., Malme, C. I. and Thomson, D. H. (1995). Marine Mammals and Noise. San Diego, CA: Academic Press.

Ridgway, S. H. and Carder, D. A. (2000). A preliminary study of loudness at frequencies of 5 to 120 kHz based on whistle response time (RT) in a dolphin. J. Acoust. Soc. Am. 108, 2515.

Risch, D., Corkeron, P. J., Ellison, W. T. and Parijs, S. M. Van (2012). Changes in humpback whale song occurrence in response to an acoustic source 200 km away. PLoS One 7, e29741.

Rolland, R. M., Parks, S. E., Hunt, K. E., Castellote, M., Corkeron, P. J., Nowacek, D. P., Wasser, S. K. and Kraus, S. D. (2012). Evidence that ship noise increases stress in right whales. Proc. Biol. Sci. 279, 2363–2368.

Roman, J., Altman, I., Dunphy-Daly, M. M., Campbell, C., Jasny, M. and Read, A. J. (2013). The Marine Mammal Protection Act at 40: status, recovery, and future of U.S. marine mammals. Ann. N. Y. Acad. Sci. 1286, 29–49.

Schecklman, S., Houser, D., Cross, M., Hernandez, D. and Siderius, M. (2011). Comparison of methods used for computing the impact of sound on the marine environment. Mar. Environ. Res. 71, 342–350.

Schick, R. S., New, L. F., Thomas, L., Costa, D. P., Hindell, M. a, McMahon, C. R., Robinson, P. W., Simmons, S. E., Thums, M., Harwood, J., et al. (2013). Estimating resource acquisition and at-sea body condition of a marine predator. J. Anim. Ecol. 82, 1300–1315.

Shannon, G., McKenna, M. F., Angeloni, L. M., Crooks, K. R., Fristrup, K. M., Brown, E., Warner, K. A., Nelson, M. D., White, C., Briggs, J., et al. (2015). A synthesis of two decades of research documenting the effects of noise on wildlife. Biol. Rev. Early view.

Simmonds, M. P. and Lopez-Jurado, L. F. (1991). Whales and the military. Nature 351, 448.

Sivle, L. D., Kvadsheim, P. H. and Ainslie, M. A. (2015a). Potential for population-level disturbance by active sonar in herring. ICES J. Mar. Sci. 72, 558–567.

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. (2015b). Severity of expert-identified behavioural responses of humpback whale, minke whale and northern bottlenose whale to naval sonar. Aquat. Mamm. 41, 469–502.

Southall, B. L., Braun, R., Gulland, F. M. D., Heard, A. D., Baird, R. W., Wilkin, S. M. and Rowles, T. K. (2006). Hawaiian melon-headed whale (*Peponacephala electra*) Mass Stranding Event of July 3-4, 2004. Silver Spring, MD: US Department of Commerce.

Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene Jr., C. R., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., et al. (2007). Marine mammal noise exposure criteria: Initial scientific recommendations. Aquat. Mamm. 33, 411–521.

Stebbins, W. C. (1966). Auditory reaction time and the derivation of equal loudness contours for the monkey. J. Exp. Anal. Behav. 9, 135–142.

Stimpert, A. K., DeRuiter, S. L., Southall, B. L., Moretti, D. J., Falcone, E. A., Goldbogen, J. A., Friedlaender, A., Schorr, G. S. and Calambokidis, J. (2014). Acoustic and foraging behavior of a Baird's beaked whale, *Berardius bairdii*, exposed to simulated sonar. Sci. Rep. 4, 7031.

Stone, C. J. (2015). Marine mammal observations during seismic surveys from 1994-2010. JNCC Report no. 463a.

Suzuki, Y. and Takeshima, H. (2004). Equal-loudness-level contours for pure tones. J Acoust Soc Am 116, 918–933.

Tasker, M. L., Amundin, M., Andre, M., Hawkins, A., Lang, W., Merck, T., Scholik-Schlomer, A., Teilmann, J., Thomsen, F., Werner, S., et al. (2010). Marine strategy framework directive: Task group 11 report: Underwater noise and other forms of energy. Luxembourg: European Union and International Council for the Exploration of the Sea.

Thomas, P. O., Reeves, R. R. and Brownell, R. L. (2015). Status of the world's baleen whales. Mar. Mammal Sci. Early view.

Tougaard, J., Carstensen, J., Teilmann, J., Skov, H. and Rasmussen, P. (2009). Pile driving zone of responsiveness extends beyond 20 km for harbor porpoises (*Phocoena phocoena* (L.)). J. Acoust. Soc. Am. 126, 11–14.

Tougaard, J., Hermannsen, L., Kyhn, L. and Polesiuk, D. (2015). Far-reaching reaction of harbor porpoises to navy mid-range sonar. In Abstracts of the 21st Biennial Conference on the Biology of Marine Mammals, San Francisco, CA.

Tyack, P. (1981). Interactions between singing Hawaiian humpback whales and conspecifics nearby. Behav. Ecol. Sociobiol. 8, 105–116.

Tyack, P., Gordon, J. and Thompson, D. (2003). Controlled-exposure experiments to determine the effects of noise on marine mammals. Mar. Technol. Soc. J. 37, 39–51.

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

Verboom, W. C. and Kastelein, R. A. (2005). Some examples of marine mammal discomfort thresholds in relation to man-made noise. In Proc. Undersea Def. Tech. Conf., Amsterdam, The Netherlands.

Verfuss, U. K., Miller, L. A. and Schnitzler, H.-U. (2005). Spatial orientation in echolocating harbour porpoises (*Phocoena phocoena*). J. Exp. Biol. 208, 3385–3394.

Visser, F., Curé, C., Kvadsheim, P. H., Lam, F.-P. A., Tyack, P. L. and Miller, P. J. O. (2016). Disturbance-specific social responses in long-finned pilot whales, *Globicephala melas*. Sci. Rep. In review.

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., Aarts, G., Sertlek, O. H., Lucke, K., Verboom, W. C., Kastelein, R. A., Ketten, D. R., van Bemmelen, R., Lam, F.-P. A., Kirkwood, R. J., et al. (2015). Assessing the impact of underwater clearance of historical explosives on harbour porpoises (*Phocoena phocoena*) in the southern North sea. Aquat. Mamm. 41, 503–523.

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, Advances in Experimental Medicine and Biology, pp. 1197–1203.

Wartzok, D., Erbe, C., Getz, W. M. and Thomas, J. (2012). Marine mammal acoustics exposure analysis models used in US Navy Environmental Impact Statements. Adv. Exp. Med. Biol. 730, 551–556.

Weir, C. R. (2008a). Overt responses of humpback whales (*Megaptera novaeangliae*), sperm whales (*Physeter macrocephalus*), and Atlantic spotted dolphins (*Stenella frontalis*) to seismic exploration off Angola. Aquat. Mamm. 43, 71–83.

Weir, C. R. (2008b). Short-finned pilot whales (*Globicephala macrorhynchus*) respond to an airgun ramp-up procedure off Gabon. Aquat. Mamm. 34, 349–354.

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. (2015). 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.

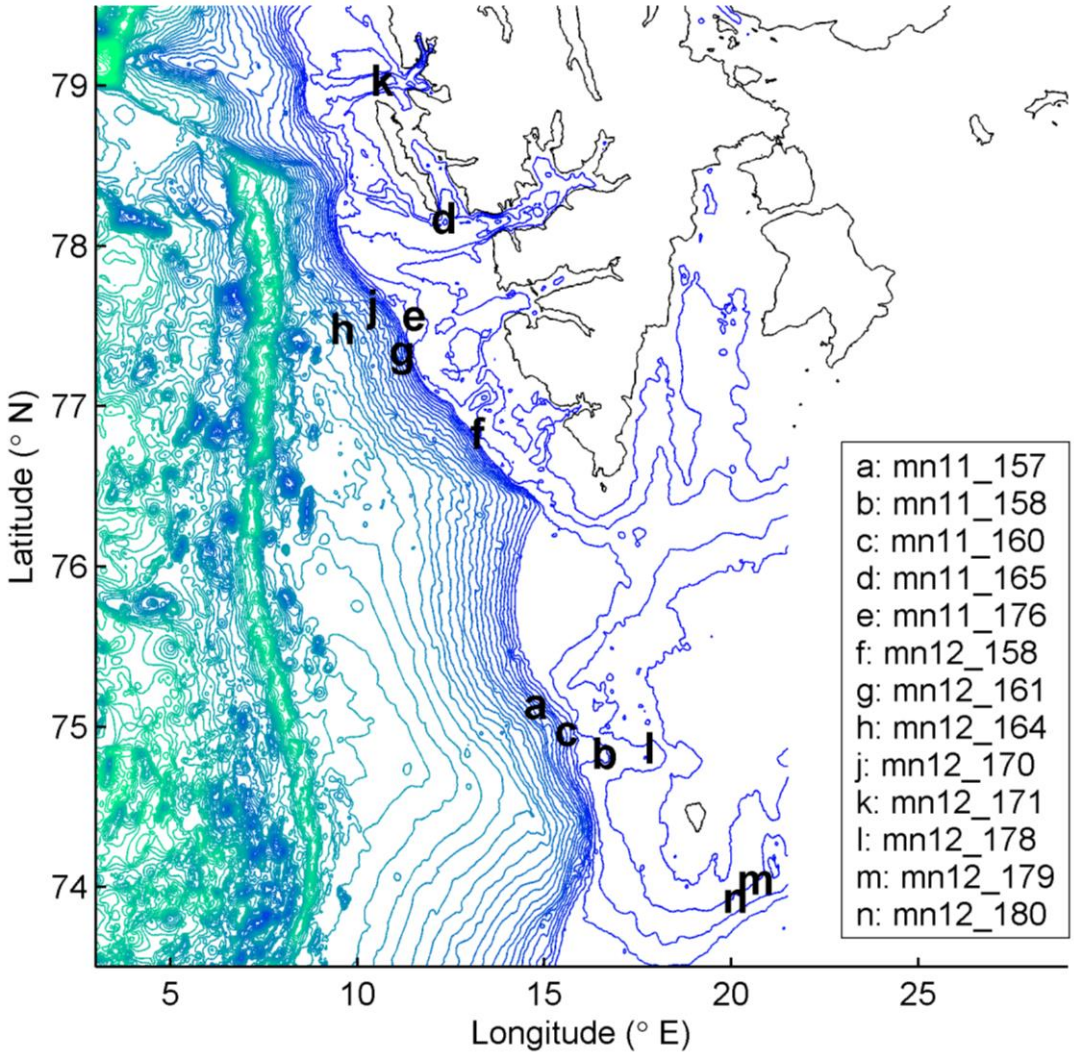
Wright, A. J., Maar, M., Mohn, C., Nabe-Nielsen, J., Siebert, U., Jensen, L. F., Baagøe, H. J. and Teilmann, J. (2013). Possible causes of a harbour porpoise mass stranding in Danish waters in 2005. PLoS One 8, e55553.

Yang, W.-C., Chou, L.-S., Jepson, P. D., Brownell, R. L., Cowan, D., Chang, P.-H., Chiou, H.-I., Yao, C.-J., Yamada, T. K., Chiu, J.-T., et al. (2008). Unusual cetacean mortality event in Taiwan, possibly linked to naval activities. Vet. Rec. 162, 184–186.

Zirbel, K., Balint, P. and Parsons, E. C. (2011). Navy sonar, cetaceans and the US Supreme Court: a review of cetacean mitigation and litigation in the US. Mar. Pollut. Bull. 63, 40–48.

FIGURES

Figure 1.1. Map of the $3S^2$ study area around Spitsbergen and Bear Island with the approximate locations of experiments conducted in 2011 and 2012. The coloured lines indicate the 100-m depth contours derived from the GEBCO_08 grid (British Oceanographic Data Centre, 2010).



Chapter 2

A method for reconstructing fine-scale tracks from dead-reckoning and position fixes applied to humpback whales

SUMMARY

Detailed information about location and movement of marine mammals is often crucial in studies on the potential effects of anthropogenic underwater noise. Combining dead-reckoning with new Fastloc-GPS technology should provide good opportunities for reconstructing georeferenced fine-scale movement tracks, and should be particularly useful for marine animals that spend most of their time under water. A computationally efficient, Bayesian state-space modelling technique was developed to estimate humpback whale locations through time in the horizontal plane. Positional observation models were based upon error measurements made during calibrations. High-resolution 3-dimensional movement tracks were produced for 13 whales using a simple process model in which errors caused by water current movements, non-location sensor errors, and other dead-reckoning errors were accumulated into a combined error term. Compared to tracks derived only from position fixes, the inclusion of dead-reckoning data greatly improved the level of detail in the reconstructed movement tracks. Using cross-validation, a clear improvement in the predictability of out-of-set Fastloc-GPS data was observed compared to more conventional track reconstruction methods. During calibration tests, Fastloc-GPS observation errors were found to vary by number of GPS satellites received and by orthogonal dimension analysed; visual observation errors varied most by distance to the whale. By systematically accounting for these observation errors in the position fixes, the developed model framework provided quantitative estimates of location uncertainty that can be appropriately incorporated into analyses of animal movement. This generic method has potential application for a wide range of marine animal species and data recording systems.

INTRODUCTION

'Dead-reckoning', i.e. predicting a location by projecting travel direction and speed from the previous location, has been used for centuries (Bowditch, 2002) and is the basis for modern inertial navigation systems in vehicles (Groves, 2013). Since dead-reckoning was introduced in bio-logging research over 25 years ago (Bramanti et al., 1988; Wilson and Wilson, 1988), it has become an established method for reconstructing fine-scale movement tracks, in particular for air-breathing marine animals that spend most of their time under water, out of sight of global positioning system (GPS) signals (Johnson et al., 2009; Wilson et al., 2007).

Dead-reckoning has been used for studies on the natural foraging and orientation behaviour of marine animals; for example for pinnipeds (e.g. Benoit-Bird et al., 2013; Davis et al., 1999; Matsumura et al., 2011; Mitani et al., 2010), turtles (Narazaki et al., 2009), diving birds (Shiomi et al., 2008; Wilson et al., 2008), and cetaceans (e.g. Aoki et al., 2012; Friedlaender et al., 2009; Goldbogen et al., 2013; Johnson et al., 2008a; Schmidt et al., 2010; Stimpert et al., 2007; Tyson et al., 2012; Ware et al., 2011; Wiley et al., 2011; Zimmer et al., 2005). Dead-reckoning has also provided important information about the behavioural responses of cetaceans to noise (Curé et al., 2015; Dunlop et al., 2013; Miller et al., 2009; Miller et al., 2012; Miller et al., 2015a; Tyack et al., 2011). Although animals can also be localised under water using active and passive sonar (e.g. Baumgartner et al., 2008; Laplanche, 2012; Mellinger et al., 2007; von Benda-Beckmann et al., 2013; Wartzok et al., 1992), such techniques require transmission and/or reception of sound which is difficult to accomplish at a high resolution, and may impact the environment of acoustically-sensitive marine mammals.

The development of miniature animal-attached data loggers that record movement parameters such as compass heading, speed, and body orientation (Burgess et al., 1998; Elkaim et al., 2006; Johnson and Tyack, 2003; Mate, 2012; Muramoto et al., 2004) enabled the use of dead-reckoning. Because each dead-reckoned position depends upon the previous one, the spatial error in the track generally grows with time due to an accumulation of sensor errors, movements of water currents, and violations of the assumptions that the animal only moves through the water in the caudo-rostral direction

and that buoyancy and lift forces are negligible (Johnson et al., 2009). A common source of uncertainty in dead-reckoning tracks is the speed of the animal. Speed may be estimated if direct measurements are missing (Miller et al., 2009), but can also be measured with a speed sensor (Wilson et al., 2008) or approximated based on pitch and change in depth (Miller et al., 2004), acoustic flow noise (Goldbogen et al., 2006), or overall dynamic body acceleration (Bidder et al., 2012).

Fixes of known positions on the earth's surface can be used to adaptively calibrate dead-reckoning sensors or to directly correct dead-reckoned positions (Groves, 2013). Position fixes of marine animals are obtained, for example, by visual observation (which can be aided by the use of laser range finders and animal-attached very high frequency (VHF) transmitters) (Miller et al., 2014; Witteveen et al., 2008), acoustic localisation (Hastie et al., 2014; Stanistreet et al., 2013), light intensity-based geolocation (Lisovski et al., 2012), or GPS satellite telemetry. Since conventional GPS is generally not feasible for marine animals because of a long (~10-30 s) time-to-fix and high current consumption (Tomkiewicz et al., 2010), new snapshot GPS technologies such as Fastloc-GPS (Bryant, 2007; Dujon et al., 2014; Schofield et al., 2007) have quickly become popular because of their ability to acquire data sufficient to estimate location during short surface intervals (Costa et al., 2012). Such approaches store GPS pseudorange data, which can be converted into positions after the logger is retrieved or after transmission through Argos (Kuhn et al., 2009) or mobile phone networks (McConnell et al., 2004). The average spatial accuracy for positions observed with Fastloc-GPS (<100 m) is much greater than for positions from Argos (0.5-10 km) or light-based geolocation (1-4°) (Costa et al., 2010; Hazel, 2009; Winship et al., 2012); therefore, the integration of Fastloc-GPS and dead-reckoning data has the potential to result in highly precise georeferenced movement tracks (Brown et al., 2013).

Most studies to date have assumed a constant bias in velocity between position fixes, essentially stretching the track to match the fixes (Wilson et al., 2007) or have iteratively approximated a constant bias (Miller et al., 2009). I describe here a new method for referencing dead-reckoning tracks to position fixes based upon state-space models (SSMs). SSMs are an appropriate statistical tool for this application because

they explicitly separate the observation processes from the underlying movement process (Patterson et al., 2008) and are a standard technique in integrated navigation systems for avian, automotive and naval applications (Groves, 2013). In animal ecology, SSMs for track reconstruction and smoothing have been implemented as Kalman filters (e.g. Anderson-Sprecher, 1994; Anderson-Sprecher and Ledolter, 1991; Johnson et al., 2008b; Nielsen et al., 2006; Patterson et al., 2010), particle filters (Breed et al., 2012), and using Markov chain Monte Marlo (MCMC) (e.g. Jonsen et al., 2005; McClintock et al., 2012; McClintock et al., 2014a; Sumner et al., 2009). Movement data of relatively low temporal resolution (e.g. collected via Argos, GPS and light-based geolocation) have been the focus of most research on marine animals, although Kalman filters have also been applied to high-resolution dead-reckoning data in combination with depth (Elkaim et al., 2006; Johnson et al., 2008a) and depth and acoustic localisation data (Ward et al., 2008).

The rapid technological developments in bio-logging will likely result in an increasing demand for analysis methods for high-resolution data that are easy to implement and fast to compute. This was achieved in the current chapter by using the fine-scale dead-reckoning track to provide the expected 2-dimensional displacement in a discrete-time correlated random walk SSM that operates at the irregular but discrete temporal scale of the low-resolution positional fixes. This has the advantage of using the high-resolution information without the computational cost associated with running a SSM at very fine temporal scale. The disadvantage is that the uncertainty associated with the dead-reckoning track is ignored, so that the estimates of uncertainty in location at times between position fixes are underestimates. The size of the underestimation depends largely on the time between position fixes, so the method will work better for animals that make frequent surfacings.

The development of the method presented in this thesis chapter was motivated by the need for detailed movement tracks of humpback whales, *Megaptera novaeangliae*. The whales were tagged with multi-sensor data loggers and Fastloc-GPS loggers in 2011 and 2012 as part of a behavioural response study (Kvadsheim et al., 2011, 2012; Sivle et al., 2015) in waters off Bear Island and Svalbard. The tagged whales were also tracked

by visual observers from a small boat. The distance between the whale and the sound source during experiments was a crucial parameter in the analyses of the behavioural responses to sonar, as errors of even a few tens of meters could have substantially changed the estimated received levels at the closest point of approach (Chapter 5). In addition, behavioural responses were more likely to be identified from higher-resolution tracks (Chapter 3). Therefore, the main objective here was to develop SSMs that integrate movement data from dead-reckoning, Fastloc-GPS, and visual observations. A secondary objective was to quantify the spatial accuracy of the Fastloc-GPS and visual (range and bearing) observations in dedicated tests, so that the observation errors included in the models would be realistic. The track reconstruction method presented here is easy to implement and has potential application for a wide range of marine animal species and data recording systems.

MATERIALS & METHODS

Data collection protocols

Thirteen humpback whales were tagged with multi-sensor digital recording tags (DTAGs, v2; Johnson et al., 2009) with a Fastloc-GPS logger (F2G 134A, Sirtrack, New Zealand) mounted on top, at northern latitudes between 74.00° and 79.03° and eastern longitudes between 9.79° and 20.68° in 2011 and 2012 (Table 2.1; Fig. 1.1). The tags were attached to each whale with suction cups using a pneumatic tag launching system (ARTS; Kvadsheim et al., 2012) or using a 15 m carbon fibre pole, cantilevered in a bow-mounted oarlock (Moore et al., 2001). The DTAGs had 1 or 2 hydrophones and recorded sound with 16-bit resolution, at 96 kHz sampling rate. The DTAGs also recorded 50 Hz pressure, temperature, tri-axial acceleration and tri-axial magnetic field-strength data. Prior to tag deployment, the internal clock of the DTAG was set to local time (synced to 1 s) using a GPS receiver. Fastloc-GPS loggers were configured to record a GPS snapshot almost instantaneously after the device emerged from the water during a surfacing of the whale. The minimum time interval between GPS snapshots was set to 30 s.

Focal follows of tagged humpback whales were conducted from an 8-m long water jet propulsion boat with an elevated observer platform. Each tag contained a very high

frequency (VHF) radio beacon which aided tracking of tagged whales. The observers on the platform measured the angle to the whale relative to the boat's heading using a protractor at the time of the animal's first surfacing observed at least 2 min after the previous sighting was recorded. Simultaneously, the (radial) line-of-sight distance to the whale was measured using a laser range finder (LRF), or estimated by eye. Because the eye height was only ~3 m, I assumed that the difference between the line-of-sight distance and the distance over the earth's surface (Lerczak and Hobbs, 1998) was negligible. To aid locating the whale at the surface, angles-of-arrival of the VHF signals from the tag were made visible to observers by a digital radio direction finder system (DFHorten, ASJ Electronic Design, Horten, Norway) connected to four 4-element Yagi antennas. All visual tracking information (e.g. range, bearing, coordinated universal time (UTC), range estimation method, and GPS positions of the observation boat at 1-s intervals) were stored in a MS Access database via the software Logger (International Fund for Animal Welfare, Yarmouth Port, MA); the data collection protocol is described in more detail elsewhere (Kvadsheim et al., 2011).

Dedicated accuracy tests

Fastloc-GPS

Dry tests with Fastloc-GPS loggers were conducted in 2011 and 2012 at four sites (56.33°N, 2.78°W; 69.68°N, 18.99°E; 78.24°N, 15.54°E; 64.92°N, 23.25°W) to quantify the spatial accuracy of each data logger. Measurements were collected with the same three loggers (device IDs: 29409, 29420, and 29510) that were deployed on humpback whales. During the calibration tests, the three devices were in a stationary position, spaced >25 cm apart, and recorded GPS snapshots every 30 s in an outdoor space with an open view of the sky. Manufacturer-provided software (Archival USB, v1.11, PathTrack, UK) was used to offload the pseudoranges and convert them into position estimates based upon the relevant daily broadcast satellite ephemeris data. Information stored for each spatial location included the UTC time stamp, number of GPS satellites used in the position calculation, and the residual value of the position solution.

For error calculations, I assumed that the true position of a logger was equal to the median of all of the observations for each logger. The geographical coordinates of the observations were converted into Universal Transverse Mercator (UTM) coordinates so that positional errors (the difference with the median coordinate) could be expressed in meters. An observation was excluded from analysis if the residual value of its position solution was >30 (no unit); this threshold was recommended by Sirtrack (Sirtrack, 2012) and adopted by other studies using Fastloc-GPS (e.g. Hoenner et al., 2012; Shimada et al., 2012; Witt et al., 2010). The error measurements were divided into bins based upon the number of satellites ('#satellite bins') from which data were recorded (4, 5, 6, 7, 8, and 9-12). Scaled *t* distributions were fitted using maximum likelihood estimation via the 'MASS' package (v7.3-19, Venables and Ripley, 2002) in the software R (v3.0.2, R Core Team, 2013) to estimate the parameters of the observation error distributions for each #satellite bin and each orthogonal dimension. The goodness-of-fit of the distributions were checked with one-sample Kolmogorov-Smirnov tests.

Visual focal follows

Five tests were conducted in June 2012 in waters near Tromsø, Norway (69.79°N, 19.19°E) and waters near Longyearbyen, Svalbard (78.56°N, 14.95°E) to quantify the accuracy of visual observations. The observers estimated range (radial distance) and bearing to an orange heavy duty inflatable buoy that had a diameter of 1.2 m. A handheld GPS receiver (Etrex Legend HCx, Garmin, Schaffhausen, Switzerland) with EGNOS capability was attached on top of the buoy for recording its GPS positions for groundtruthing. A total of seven observers participated in the tests (the same individuals who conducted the focal follows on tagged humpback whales); two or three observers participated at the same time. The observation boat from where visual estimates were made sailed an undetermined course, making occasional turns, matching operations during whale tracking. To imitate the data coverage during real focal follows, the boat was within <200 m from the buoy for roughly 50% of the estimates but occasionally moved to distances of around 1 km. One person (the 'data recorder') stored the estimates in the software Logger and gave vocal commands. Once every 2 min, the data recorder called out "Ready", which indicated to the observers to start looking for the target and to the driver to adopt a steady course. About 10 seconds later, the data

recorder called out “Mark”, which indicated to the observers to make their estimates and write them down on paper. I limited the time that the observers could look at the target because this influences the accuracy of the range estimates (Williams et al., 2007). The estimates for range were made visually by the observers, and protractors were used to measure the bearing relative to the heading of the boat. The same observation boat and data collection protocol were used during the focal follows of the humpback whales (details in Kvadsheim et al., 2011).

The absolute bearing (relative to true north) to the whale from the boat at the time of a sighting was calculated by adding the boat’s course-over-ground derived from GPS to the relative bearing to the whale. Linear errors in range and bearing were calculated as the difference between the visual estimates and the ‘true values’ derived from the GPS positions of the buoy and the observation boat. The linear range errors were clearly a function of range itself (and thus ‘heteroskedastic’), so percent error in range was used instead of absolute error (i.e. a multiplicative error model was used). To test for potential remaining range-dependency, I fitted a linear regression model to the percent error in range as function of true range in MATLAB (v8.1; The Mathworks, Natick, MA). A wrapped Cauchy distribution was fitted to the angular errors in bearing in R using the package ‘circular’ (v0.4-7, Agostinelli and Lund, 2013).

Process model

Position fixes (with respect to the Earth frame of reference) of the humpback whale at the sea surface naturally occurred at irregular time intervals. The process model in the model framework operated on the relatively coarse time scale of these fixes. This greatly reduced computational time, but had the disadvantage that the dead-reckoning errors were not fully incorporated and thus underestimated the positional uncertainty between fixes. The SSM described here is therefore an approximation to a full SSM that would run on the finer time scale of the tag data. The humpback whale data set contained relatively high rates of position fixes (average of 0.1-1.9 observations/min; n=13; Table 2.1), and at those rates the contribution of dead-reckoning on the uncertainty was relatively minor compared to the uncertainty from the positional

observations. I therefore combined a fairly simple process model with relatively realistic positional observation models.

For the process model, I defined J as the number of position fixes, j=1,…,J as the index over these fixes, and Δ_j as the time interval between t_j and t_{j+1}. Scalars were written in italic and vectors in bold italic. Only the horizontal (*xy*) plane was considered because the depth of the whale (i.e. the *z*-coordinate of its position) was measured with a highly accurate sensor and therefore assumed to be observed without error. The process model essentially combined the whale's position given by the high-resolution dead-reckoning track (see next section) with a velocity correction term. Specifically, given an initial unobserved whale position $\boldsymbol{x}_1$, the unobserved whale positions $\boldsymbol{x}_j$ at t_j were derived using the algorithm

$$\boldsymbol{x}_{j+1} = \boldsymbol{x}_j + \boldsymbol{d}_j^{dr} + \boldsymbol{v}_j^{cor}\Delta_j, \qquad (2.1)$$

where $\boldsymbol{d}_j^{dr}$ is the whale's expected displacement over Δ_j given by the uncorrected dead-reckoning track, and $\boldsymbol{v}_j^{cor}$ is the velocity correction for the track segment. This correction term can be interpreted as the mean 'bias' or 'drift' in velocity over Δ_j (Bowditch, 2002; Wilson et al., 2007), although in many studies using movement models these qualifications refer to the mean velocity of the animal itself (McClintock et al., 2014b). To reflect the belief that $\boldsymbol{v}^{cor}$ could only change slowly over time, I assumed that its process was a non-directional first-order Gaussian random walk,

$$\boldsymbol{v}_{j+1}^{cor} \sim MVN(\boldsymbol{v}_j^{cor}, \boldsymbol{\Sigma}\Delta_j), \qquad (2.2)$$

where the process noise variance-covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$ and σ_x^2 and σ_y^2 represent the variances for the *x*- and *y*-dimension. The covariance term was set to 0 as the process noise was assumed to be independent between the two spatial dimensions. A linear relationship of $\boldsymbol{\Sigma}$ with Δ_j was incorporated to account for the dead-reckoning errors that grow with time.

Determining the dead-reckoning track

I describe here how the uncorrected dead-reckoning track was derived from the high-resolution observations. As mentioned earlier, no observation models were incorporated for these tag-derived data. Parameter *I* was defined as the number of high-resolution observations, *i*=1,…,*I* as the index over these observations, and Δ_i as the time interval between t_i and t_{i+1}. The whale's uncorrected velocity $\boldsymbol{v}_i$ for Δ_i was

$$\boldsymbol{v}_i = s_i \cos(p_i) \begin{bmatrix} \cos(h_i) \\ \sin(h_i) \end{bmatrix}, \qquad (2.3)$$

where s_i is the whale's speed-through-water, and pitch p_i and heading h_i describe the orientation of the whale's body with reference to the Earth frame (Johnson and Tyack, 2003). Vector $\boldsymbol{v}_i$ may be used to calculate the uncorrected dead-reckoning track using the algorithm $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \boldsymbol{v}_i \Delta_i$; however, because the process model operated on the coarser, irregular time scale t_j determined by the position fixes, I integrated $\boldsymbol{v}_i$ with respect to time in the domain $t_i = [t_j, t_{j+1})$ to find the whale's uncorrected displacement $\boldsymbol{d}_j^{dr}$ that was used in Eqn 2.1:

$$\boldsymbol{d}_j^{dr} = \sum_{t_i=t_j}^{t_{j+1}} (\boldsymbol{v}_i \Delta_i). \qquad (2.4)$$

Positional observation models

A set of equations stochastically related each whale's unobserved position $\boldsymbol{x}_j$ at time t_j to the observations of range (radial distance), bearing, and/or Fastloc-GPS. The observation error structures were based upon the results of the dedicated accuracy tests (see 'Results'). Specifically, the observation model relating the observed Fastloc-GPS position, $X_{x,j}^F$, to the unobserved whale position for the *x*-dimension was

$$X_{x,j}^F \sim t\left(x_{x,j}, \sigma_{x,q}^F, \upsilon_{x,q}^F\right), \qquad (2.5)$$

with a similar formulation for the *y*-dimension. Parameter $\boldsymbol{\sigma}_q^F$ represents the scale and $\boldsymbol{\upsilon}_q^F$ the shape (or, degrees of freedom) of the scaled *t* distribution. Because Fastloc-GPS

accuracy is related to #satellites (Dujon et al., 2014; Hazel, 2009), I used the parameter estimates obtained from the dry test data as fixed values for $\boldsymbol{\sigma}_q^F$ and $\boldsymbol{v}_q^F$ (where quality q=1,…,6 indexes the 4, 5, 6, 7, 8, and >8 satellite bins, respectively) in an approach similar to the use of Argos quality classes in other studies (e.g. Jonsen et al., 2005).

The observation model implemented for range between observer and whale at the surface was

$$R_j \sim N\left(r_j, r_j \sigma_m^r / 100\right), \qquad (2.6)$$

where R_j is the observed range and r_j is the unobserved range. Thus, the observation error was assumed to be normally-distributed around 0%, which was close to the truth according to the visual observer tests (see 'Results'). Scale parameter σ_m^r represents the percent error SD for m = 1, 2, where range estimation method m = 1 if estimates were made visually (by eye), and m = 2 if a laser range finder was used to make the measurement. Its value for m = 1 was based upon the visual accuracy tests and for m = 2 was assumed to be 10%. The observation model implemented for absolute bearing between the observer and the whale was

$$\Phi_j \sim wC(\varphi_j, \rho), \qquad (2.7)$$

where Φ_j is the observed bearing, φ_j the unobserved bearing, and ρ is the scale (or, concentration) of the wrapped Cauchy distribution that was derived from the visual accuracy tests.

Finally, I related the unobserved difference in position between the observation boat and the whale ($\boldsymbol{d}_j^{bw} = \boldsymbol{x}_j - \boldsymbol{x}_j^b$) to the unobserved range and bearing via a Cartesian-to-polar coordinate transformation:

$$r_j = \left\|\boldsymbol{d}_j^{bw}\right\|, \text{ and} \qquad (2.8)$$

$$\varphi_j = \tan^{-1}(d_{x,j}^{bw}/d_{y,j}^{bw}), \qquad (2.9)$$

where $\tan^{-1}$ is the four-quadrant arctangent to realise φ_j=(−180°, 180°]. The position of the observation boat $\boldsymbol{x}_j^b$ was measured with a GPS receiver with an average error of <3 m (unpublished data). This GPS receiver was located within 1 m from the visual observers; therefore, $\boldsymbol{x}_j^b$ was set to be equal to the Cartesian coordinates of the measured GPS positions (the model can be easily adapted to include error on the observer boat's position).

Data processing and model fitting

Pre-processing

Procedures for offload, calculation and filtering of data collected by the deployed Fastloc-GPS loggers were the same as for test data (see for details: 'Methods – Dedicated accuracy tests'). Using a conversion from geographical to UTM coordinates, all positions of the whale and the observation boat were placed in a Cartesian coordinate system with at the origin (x=0, y=0) the first observed position of the whale (Table 2.1). I temporally aligned the position fixes of the same surfacing to further reduce computational costs. This was accomplished by 1) identifying pairs of Fastloc-GPS observations that were observed within 5 s of one another and replacing the timestamp of the last fix with that of the first (only for whales that had two GPS loggers attached), and 2) replacing the timestamps of the visual observations that were made ±5 s from a Fastloc-GPS observation by the timestamp of the Fastloc-GPS observation. The 5-s interval was judged to be the longest time interval that could not result in observations from separate whale surfacings being falsely aligned, and was based upon an exploratory analysis in which the times of position fixes were plotted on the corresponding dive profile.

For each tag record, data on depth, acceleration and magnetic field strength from the DTAG were downsampled to 1 Hz resolution (Δ_i = 1 s) using a DC accurate decimating filter. The whale's pitch (p_i) and heading (h_i) were derived from the acceleration and magnetic field measurements following the techniques detailed elsewhere (Johnson and Tyack, 2003). Estimates of the whale's speed-through-water (s_i) based upon depth rate

per second divided by the sine of pitch during steep (i.e. $|p_i|$>50°) descents and ascents (Miller et al., 2004) were regressed against the uncalibrated (1-s root-mean square) noise level (L_i) in the 66-94 Hz frequency band (Ware et al., 2011) using the model:

$$\log(s_i) \sim N(\beta_0 + \beta_1 L_i, \sigma^L), \qquad (2.10)$$

where β_0, β_1 and σ^L are model parameters. This function should be an appropriate model according to the physics of flow noise (Haddle and Skudrzyk, 1969), although empirical verification is recommended on a case-by-case basis. Both body pitch and noise level were low-pass filtered using a zero-group-delay fast impulse response filter with a 0.15 Hz cut-off frequency to remove fine-scale temporal variations such as from fluke strokes to generate thrust (Simon et al., 2012). The fitted function was used to predict s_i from L_i throughout the entire tag record, including the regions of shallower pitch (Goldbogen et al., 2006; Goldbogen et al., 2008). Flow noise is likely to be influenced by noise generated by the sea surface when the whale is at shallow depth; therefore, speed-through-water estimates for each period where the whale was at <5 m depth were replaced using a linear interpolation of the start and end values of the period.

Fitting the track reconstruction model

Model fitting was performed using Markov chain Monte Carlo (MCMC) algorithms in the software JAGS (v3.4.0, Plummer, 2003) through an interface with MATLAB. Uniform priors were assigned to most parameters: $\sigma_x \sim Unif(0, 0.1)$, $\sigma_y \sim Unif(0, 0.1)$, $v_{x,1}^{cor} \sim Unif(-1, 1)$, and $v_{y,1}^{cor} \sim Unif(-1, 1)$; only the initial position of the whale had informative priors that reflected the accuracy of its observation (Table 2.2). Thirteen models were fitted to the data set; one for each whale record. To assess whether parameters converged to stationary distributions, I ran two MCMC chains with different initial values. Each chain had a burn-in period of 200,000 samples and a total run length of 280,000 samples, and was downsampled (thinned) by a factor of 5 to reduce memory load. Mixing and stationarity were assessed by visual examination of trace plots and using the Brooks-Gelman-Rubin statistic $\hat{R}$ (Gelman et al., 1998). MCMC chains were run in parallel on multiple cores of a desktop computer (Intel i7-4930K processor with

six physical cores; 32 GB of RAM; 64-bit MS Windows 7 operating system); up to three models were fitted at the same time.

Post-processing

The JAGS output included the posterior estimates of the low-resolution track ($\boldsymbol{x}_j$; the whale positions at the times of the position fixes); posterior estimates of the high-resolution track ($\boldsymbol{x}_i$) were calculated in a post-processing analysis. To obtain the final (corrected) position estimates with uncertainty, 3,200 high-resolution track realisations (or 'posterior sample tracks') were calculated from 1,600 computed iterations (10% of the total) using the whale's uncorrected velocity $\boldsymbol{v}_i$ derived with Eqn 2.3, the posterior samples of the whale's initial position $\boldsymbol{x}_1$, and the posterior samples of the velocity correction $\boldsymbol{v}_j^{cor}$.

Assessing model performance

A form of 10-fold cross-validation (Fielding and Bell, 1997) was conducted to compare the performance of the developed method to other track reconstruction methods. Specifically, the cross-validation analyses tested how well out-of-set Fastloc-GPS positions were predicted by the state-space model and the other methods. Only Fastloc-GPS position fixes were part of this analysis as they were generally more accurate than the visual position fixes (see 'Results') and less likely to include temporal autocorrelation. First, I left out every 10th Fastloc-GPS observation (the 'validation data') and fitted the state-space model to the remaining observations (the 'training data'). For each observation in the validation set, I then measured the positional (cross-validation) error relative to the following horizontal track types: 1) the mean posterior track based on the state-space model fitted to the training data, 2) a track with linear interpolation between the training data, 3) a track with linear interpolation between visual position fixes (excluding fixes that occurred during the same surfacings as the validation data), and 'forced-point' dead-reckoning tracks that were stretched to match the training data (Wilson et al., 2007) and initially calculated with 4) constant speed or 5) speed derived from flow noise. The procedure was iterated 10 times per whale, each time changing the validation set indices to leave out a different 10% of the Fastloc-GPS observations. Cross-validation analyses were conducted for three different whales (IDs

1, 7, and 11) and positional errors were averaged within their #satellite bin to assess overall model performance.

Because the rate of Fastloc-GPS fixes was relatively high for these three whales (~1 fix every 2 min; Table 2.1), a second type of cross-validation was conducted in which the validation set was created by taking a series of five consecutive positions instead of a single position, leaving the next five consecutive positions in the training data set. Therefore, instead of omitting 10% of the observations at each iteration, 50% of the observations were omitted (periods that averaged 10 min) at each iteration, and the same Fastloc-GPS positions were part of the validation set five times. Calculation of the positional cross-validation errors was the same as described above, except that visual position fixes were excluded during the whole time interval spanning the five consecutive Fastloc-GPS observations.

RESULTS

Fastloc-GPS accuracy tests

A total of 35,347 location observations were collected during 'dry' tests with Fastloc-GPS loggers (n=3) in fixed positions, which amounted to a total of 4.9 days' worth of data. The number of observations assigned to the 4, 5, 6, 7, 8, >8 satellite bins was 3,864 (11%), 4,690 (13%), 5,648 (16%), 6,402 (18%), 6,102 (17%) and 8,641 (24%), respectively. Only 0.2% of these observations had residual values >30 and were omitted from the final data set (all sites and devices combined). The spatial errors of the three loggers were similar within each #satellite bin, although one logger (ID 29420) acquired data from a greater number of satellites on average (7.7) than the other two loggers (6.5 and 6.7) (Fig. 2.1) and thus recorded more positions of higher accuracy. There were some indications that the errors differed somewhat across test sites, possibly because of differential weather conditions, but this comparison was limited by low numbers of observations in some of the subsets (Fig. 2.1). For both spatial dimensions (x and y), the accuracy of the Fastloc-GPS observations was positively related to the #satellites used in the position calculation (Fig. 2.2). The positional errors in the final data set were well described by the scaled t distribution (Fig. 2.2; Kolmogorov-Smirnov tests, $p>0.05$ for each distribution). The maximum likelihood estimates and standard errors (SEs) for $\boldsymbol{\mu}^F$,

$\boldsymbol{\sigma}^F$, and $\boldsymbol{\upsilon}^F$ are provided in Table 2.3. The obtained error distributions were symmetric (μ^F close to 0 m) and ~1.3 times narrower in the *x*-direction than in the *y*-direction (σ_y^F/σ_x^F; see also Figs. 2.1 and 2.2). Estimates for $\boldsymbol{\upsilon}^F$ increased with the #satellites from about one (Cauchy errors) for 4 satellites to about eight (approximating Gaussian errors) for >8 satellites.

Visual accuracy tests

The accuracy tests with human observers (n=7) produced a total of 220 visual observations of range and bearing used to estimate location. Each test took ~40 min; the combined duration of the data collection periods was 3.2 h. Despite modest sample sizes, the percent errors in range and angular errors in bearing were reasonably well described by the Normal and wrapped Cauchy distributions, respectively (Fig. 2.3; Kolmogorov-Smirnov test with range data, p>0.05). The slope of the percent error in range regressed against the true range was significantly different from 0 at p=0.02, indicating that the percent error overestimated at close range and underestimated at large range, but this effect was very small (0.027% per metre; Fig. 2.3). There was very little consistent negative bias in the estimates of range (μ: −2.95%) and bearing (μ: −1.24°). Visual estimates of range were relatively inaccurate (σ_1^r: 30.2%) compared to the bearing estimates (ρ: 0.897; circular SD: 11.6°). The positional uncertainty of a whale location obtained through visual observation will therefore be highly asymmetrical in Cartesian coordinates, further justifying the use of a range-and-bearing observation model to incorporate the anisotropic errors.

Humpback whale tracks

Visual examination of the trace plots of the estimated parameters confirmed that convergence was always reached within the burn-in phase, MCMC chains were stationary, and sufficient posterior samples were obtained. This was corroborated by $\hat{R}$ values of ≤1.05 for each parameter (Table 2.4). The model runtime varied greatly across whales (range: 0.2 to 78.4 h; Table 2.1) and depended strongly upon the number of position fixes (especially from Fastloc-GPS).

I first provide an example of a reconstructed fine-scale track using the results for whale 11. This whale remained in an area of about 5 × 4 km ($x \times y$) for the full 7.8 h duration of the track (Fig. 2.4). The whale's horizontal movements ranged from very directional with slow clockwise turns and little short-term heading variation to very non-directional with large short-term heading variation. In general (and as expected), the most probable (posterior mean) whale positions were very close to the Fastloc-GPS fixes, further from position fixes made with laser range finder, and the furthest from position fixes for which range was estimated by eye (Fig. 2.4). Repetitions of bursts of speed (up to 3-4 ms^{-1}) concordant with rapid changes in depth suggested that this whale performed multiple feeding 'lunges' (i.e. feeding events in which the animal speeds up to engulf large volumes of water and filter prey; Goldbogen et al., 2008; Simon et al., 2012) in the bottom phase of most dives. The whale's uncorrected velocity $\boldsymbol{v}$ over the whole track ranged between −3.6 and 3.2 ms^{-1} in the x-direction (min/max v_x^{cor}: −0.2/1.0 ms^{-1}) and between −4.3 and 4.0 ms^{-1} in the y-direction (min/max v_y^{cor}: −0.9/0.3 ms^{-1}). Some sudden changes in $\boldsymbol{v}^{cor}$ appeared to correspond with changes in the movement parameter values for this animal (e.g. the shallow diving period starting at 04:00 UTC). The velocity correction process for this whale was relatively volatile (posterior means for σ_x and σ_y of 0.014 and 0.012 ms^{-1}, respectively) compared to that of other whales (Table 2.4; Appendix I).

The complete data set of 13 whales contained large differences in movement patterns and behaviour (Appendix I), and detailed visual inspection of the tracks suggested that the track reconstruction model performed satisfactory under a wide range of conditions. The positional uncertainty in tracks with none or few Fastloc-GPS fixes (e.g. whales 2 and 3; Table 2.1) was generally greater than for tracks with many Fastloc-GPS fixes (e.g. whales 7 and 13). Clear differences in the posterior mean estimates of $\boldsymbol{v}^{cor}$ were observed among animals (Appendix I); while in some cases its values remained close to 0 ms^{-1} for the entire track duration (e.g. whales 1 and 9), in others its values gradually changed over time (e.g. whale 13) or values indicated a strong consistent bias in one direction (whale 3). This between-animal variation in $\boldsymbol{v}^{cor}$ was also reflected in the posterior mean estimates of $\boldsymbol{\sigma}$, which ranged between 0.003 and 0.015 ms^{-1} and were often similar between x- and y-dimensions (Table 2.4; Fig. 2.5).

Cross-validation analyses

Results of the cross-validations were based upon a combined (n=3 whales) validation set of 206, 247, 212, 161, 96, 44, and 29 unique Fastloc-GPS positions (for 4, 5, 6, 7, 8, 9, and >9 satellites, respectively). Positional cross-validation errors indicated that the mean posterior tracks of the Bayesian SSMs most closely approximated the validation data and the mean measurement errors from the dry tests compared to other track reconstruction methods (Fig. 2.6). Performance varied across methods, with the forced-point dead-reckoning tracks being, on average, more accurate than the tracks with linear interpolation between Fastloc-GPS fixes and tracks with linear interpolation between visual fixes (Fig. 2.6). Mean cross-validation errors decreased with increasing #satellites for all track types, indicating that the measurement errors of the validation data formed part of the cross-validation errors. As expected, the cross-validation errors were greater and the differences between methods greater when the validation sets contained blocks of 5 consecutively observations (simulating periods of ~10 min without data collection) instead of single observations (Fig. 2.6). However, the above results regarding which method performed best and the decreasing error with #satellites were the same for both 10% and 50% data removal.

DISCUSSION

Accurate tracking of marine animals (e.g. mammals, penguins, and turtles) with high-resolution multi-sensor data loggers has become increasingly important in ecology and conservation biology (Johnson et al., 2009; Nathan et al., 2008). These data loggers have already provided valuable information on topics such as foraging behaviour (Aoki et al., 2015; Goldbogen et al., 2008; Parks et al., 2014; Samarra and Miller, 2015; Wiley et al., 2011), time and energy budgets (Friedlaender et al., 2013; Isojunno and Miller, 2015) and human impacts (Antunes et al., 2014; Miller et al., 2015a; Wensveen et al., 2015b), but the number of methods available for analysis of marine animal movements from high-resolution data is still very limited. To partially address this gap, this chapter describes an effective SSM framework that is designed for relatively fast reconstruction of fine-scale tracks combining visual, Fastloc-GPS, and dead-reckoning data. Empirical data from accuracy tests formed the basis of the observation models.

Visual observation is a method that is often used for accurate tracking of marine mammals at the surface (e.g. using land-based theodolite tracking: Dunlop et al., 2013; boat-based focal follows: Miller et al., 2014; or stereo photogrammetry: Macfarlane et al., 2015), but a quantitative assessment of its accuracy, as presented in this study, is relatively uncommon. The visual accuracy tests with a floating buoy showed that the errors in range generally contributed most to the combined positional error from range and bearing observations, which is consistent with results from more extensive testing during transect line surveys (Leaper et al., 2010). The average range estimation error (SD: 30%) was similar to those of naturalists on whale-watching vessels (25%) and less similar to range estimates of captains (19%) and members of the general public (45%) on these same vessels (Baird and Burkhart, 2000).

The use of the normally-distributed percent error for range was a practical way to scale the error with distance, although a minor range-dependent effect in the transformed data remained. Error models for range based upon distributions such as the gamma or log-normal may be more appropriate in certain situations (Marques, 2004). The accuracy tests were designed to emulate the real focal follows as much as possible by, for example, using the same platform and observers, and limiting the duration that the target was visible to the observers (Williams et al., 2007). However, these tests were not exhaustive and the estimated errors were likely only reasonable approximations to the actual errors during focal follows. Observer-specific differences in the visual estimates were not modelled for a number of reasons (i.e. recording who made each observation was not part of the field protocol, low sample size per observer for accuracy tests, and the estimation error of one observer from 2011 was not quantified), but future studies could incorporate observer-specific range and bearing errors within the model framework.

The estimated accuracy of the three Fastloc-GPS loggers was roughly comparable to other reports (Bryant, 2007; Dujon et al., 2014; Hazel, 2009) when accuracy was quantified in terms of 1-dimensional spatial error (Fig. 2.7). For example, I found that 50 and 95% of the errors in positions based on 4 GPS satellites were within 50 and 633

m, respectively. In comparison, the values for these respective percentiles in Bryant (2007) were 50 and 810 m and in Dujon and colleagues (2014) were 36 and 724 m. The differences in accuracy compared to these other studies were likely caused by factors related to satellite coverage, atmospheric conditions and individual receiver sensitivity. One important conclusion from the calibration tests was that Fastloc-GPS errors differed between the two orthogonal dimensions, as has been described for the Argos system (Vincent et al., 2002). It is therefore advisable to always report the latitude/northing error and longitude/easting error separately.

The on-animal accuracy of Fastloc-GPS loggers may vary somewhat from the accuracy measured during dry tests because of variation in tag placement position on the animal, recording settings, and slowly-changing atmospheric effects such as humidity, pressure, and ionospheric delay. Therefore, in the future, such covariates could be incorporated within SSM frameworks to investigate their relative contributions or to further improve measurement error structures and track accuracy.

Visual tracking and Fastloc-GPS are relatively accurate compared to most alternative positioning technologies (such as Argos; Patterson et al., 2010; Witt et al., 2010), and some research questions can be sufficiently addressed without the use of complex methods such SSMs. Possible alternatives to SSMs are removing part of the data based upon unrealistic speeds (McConnell et al., 1992) or based upon the number of satellites used in the position calculation (Hazel, 2009; Shimada et al., 2012). Also, various interpolation methods are available for estimating the track between known position fixes (Tremblay et al., 2006).

There are many sources of error that can influence dead-reckoning of animals under water. Eqn 2.3 hints at one such source of error; the animals naturally move in the water frame of reference and speed is measured in this frame, but the orientation of the whale, used to derive velocity, is measured in the Earth frame (which is eventually of most interest). In addition, water currents may vary with depth due to the Ekman spiral, sensor errors accumulate with time, and speed estimates are often biased and not continuously observed. Also, marine animals do not always move in the same direction

as their (flexible) body is oriented due to inertia, buoyancy, and hydrodynamic lift forces (caused by large pectoral fins, for example) (Johnson et al., 2009). Suction-cup tags can occasionally move over the whale's body, which means that the correction angles for the conversion from tag to animal frame, as well as the flow noise/speed-relationships, may vary throughout the tag record. Because of this complex mix of errors, I essentially sacrificed some realism for practicality and implemented a relatively simple process model as a correlated random walk on the joint error in horizontal velocity. Visual inspection indicated that $\boldsymbol{v}^{cor}$ co-varied with the movement parameters for some animals, but in other tracks small and consistent offsets likely caused by water current appeared to be the dominant factor (Appendix I). More in-depth analysis of the estimates of $\boldsymbol{v}^{cor}$ may provide further insights in the relative contributions of the sources of errors in the tracks.

Fitting the models with MCMC had the advantage that the non-Gaussian observation error structures for Fastloc-GPS and bearings were easy to implement, but also made model fitting relatively slow (Table 2.1). To make model fitting with MCMC possible, measurement errors were not modelled at the time step of the high-resolution data. As a result, the model underestimated the positional uncertainty in the track when fixes were not observed. This effect was likely to be small for the short track segments in this study but will increase with the time since the most recent location measurement. More realistic confidence bounds could conceivably be added to the track segments between surfacings using a Kalman filter that is conditioned on the start and end points of each track realisation.

By accounting for the observation errors in the position fixes, the model can provide a clear improvement over simpler method to georeference dead-reckoning tracks (Wilson et al., 2007). Similarly, compared to tracks derived only from position fixes (Sivle et al., 2015b), the inclusion of dead-reckoning data greatly improved the level of detail in the reconstructed humpback whale tracks (Appendix I). Cross-validation analyses confirmed that out-of-set Fastloc-GPS locations were better predicted by the model framework presented here than by simpler track reconstruction methods that do not allow for positional observation error. Independent validation of the technique presented

might be (partially) possible in the future using double tagging experiments (e.g. Winship et al., 2012) with conventional GPS, using passive acoustic locations of animals that vocalise underwater (Laplanche et al., 2015), or using current velocity data from acoustic Doppler current profilers or numerical ocean models.

Being a recursive method, dead-reckoning generally results in positional errors that increase with time, and the speed of the water current may have a particularly large influence on these errors. Knowing the rate at which model performance deteriorated would be useful for scientists studying different species or for users of animal data loggers who need to decide on position sampling schemes. However, a preliminary analysis (not shown here) of the cross-validation errors against time to the nearest Fastloc-GPS position did not consistently demonstrate this trend of decreasing model performance, likely because of the relatively large contribution of Fastloc-GPS observation errors and because time intervals between locations were relatively short (<10 min).

The integration of Fastloc-GPS, depth, speed and inertial sensor data is an exciting development that opens the door to the reconstruction of georeferenced 3-dimensional movement tracks with relatively high precision compared to existing positioning methods. As similar track reconstruction approaches are currently being developed (Battaile et al., 2015; Laplanche et al., 2015; Zidek and Trites, 2014), a systematic comparison of the tracks produced by the different techniques in the future would be valuable. High-resolution animal tracks have the potential to answer fascinating scientific questions about, for example, predator movements in relation to prey fields, dynamics of group movement, impacts of disturbance on fine-scale animal behaviour, and how foraging effort and success relate to individual and population fitness. The advancement of bio-logging technology is rapid and, in my opinion, scientists will benefit from the use and development of analysis methods that make the most out of the growing wealth of information.

REFERENCES

Agostinelli, C. and Lund, U. (2013). R package “circular”: circular statistics.

Anderson-Sprecher, R. (1994). Robust estimates of wildlife location using telemetry data. Biometrics 50, 406–416.

Anderson-Sprecher, R. and Ledolter, J. (1991). State-space analysis of wildlife telemetry data. J. Am. Stat. Assoc. 86, 596–602.

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

Aoki, K., Amano, M., Mori, K., Kourogi, A., Kubodera, T. and Miyazaki, N. (2012). Active hunting by deep-diving sperm whales: 3D dive profiles and maneuvers during bursts of speed. Mar. Ecol. Prog. Ser. 444, 289–301.

Aoki, K., Amano, M., Kubodera, T., Mori, K., Okamoto, R. and Sato, K. (2015). Visual and behavioral evidence indicates active hunting by sperm whales. Mar. Ecol. Prog. Ser. 523, 233–241.

Baird, R. W. and Burkhart, S. M. (2000). Bias and variabiliy in distance estimation on the water: implications for the management of whale watching. IWC Meeting Document SC/52/WW1.

Battaile, B. C., Nordstrom, C. A., Liebsch, N. and Trites, A. W. (2015). Foraging a new trail with northern fur seals (*Callorhinus ursinus*): Lactating seals from islands with contrasting population dynamics have different foraging strategies, and forage at scales previously unrecognised by GPS interpolated dive data. Mar. Mamm. Sci. 31, 1494–1520.

Baumgartner, M. F., Freitag, L., Partan, J., Ball, K. R. and Prada, K. E. (2008). Tracking large marine predators in three dimensions: the real-time acoustic tracking system. IEEE J. Ocean. Eng. 33, 146–157.

Benoit-Bird, K., Battaile, B., Nordstrom, C. and Trites, A. (2013). Foraging behavior of northern fur seals closely matches the hierarchical patch scales of prey. Mar. Ecol. Prog. Ser. 479, 283–302.

Bidder, O. R., Soresina, M., Shepard, E. L. C., Halsey, L. G., Quintana, F., Gómez-Laich, A. and Wilson, R. P. (2012). The need for speed: testing acceleration for

estimating animal travel rates in terrestrial dead-reckoning systems. Zoology (Jena) 115, 58–64.

Bowditch, N. (2002). The American Practical Navigator. Bethesda, MD: National Imagery and Mapping Agency.

Bramanti, M., Dall'Antonia, L. and Papi, F. (1988). A new technique to monitor the flight paths of birds. J. Exp. Biol. 134, 467–472.

Breed, G. A., Costa, D. P., Jonsen, I. D., Robinson, P. W. and Mills-Flemming, J. (2012). State-space methods for more completely capturing behavioral dynamics from animal tracks. Ecol. Modell. 235-236, 49–58.

Brown, D. D., Kays, R., Wikelski, M., Wilson, R. and Klimley, A. P. (2013). Observing the unwatchable through acceleration logging of animal behavior. Anim. Biotelemetry 1, 20.

Bryant, E. (2007). 2D Location Accuracy Statistics for Fastloc® Cores Running Firmware Versions 2.2 & 2.3. Wildtrack Telemetry Systems Ltd.

Burgess, W. C., Tyack, P. L., Le Boeuf, B. J. and Costa, D. P. (1998). A programmable acoustic recording tag and first results from free-ranging northern elephant seals. Deep. Res II 45, 1327–1351.

Costa, D. P., Robinson, P. W., Arnould, J. P., Harrison, A. L., Simmons, S. E., Hassrick, J. L., Hoskins, A. J., Kirkman, S. P., Oosthuizen, H., Villegas-Amtmann, S., et al. (2010). Accuracy of ARGOS locations of Pinnipeds at-sea estimated using Fastloc GPS. PLoS One 5, e8677.

Costa, D. P., Breed, G. A. and Robinson, P. W. (2012). New insights into pelagic migrations: implications for ecology and conservation. Annu. Rev. Ecol. Evol. Syst. 43, 73–96.

Curé, C., Doksaeter 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.

Davis, R. W., Fuiman, L. A., Williams, T. M., Collier, S. O., Hagey, W. P., Kanatous, S. B., Kohin, S. and Horning, M. (1999). Hunting behavior of a marine mammal beneath the Antarctic fast ice. Science 283, 993–996.

Dujon, A. M., Lindstrom, R. T. and Hays, G. C. (2014). The accuracy of Fastloc-GPS locations and implications for animal tracking. Methods Ecol. Evol. 5, 1162–1169.

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.

Elkaim, G. H., Decker, E. B., Oliver, G. and Wright, B. (2006). Marine Mammal Marker (MAMMARK) dead reckoning sensor for in-situ environmental monitoring. In Proceedings of the IEEE/ION Position, Location, And Navigation Symposium, pp. 976–987.

Fielding, A. H. and Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence / absence models. Environ. Conserv. 24, 38–49.

Friedlaender, A. S., Hazen, E. L., Nowacek, D. P., Halpin, P. N., Ware, C., Weinrich, M. T., Hurst, T. and Wiley, D. (2009). Diel changes in humpback whale *Megaptera novaeangliae* feeding behavior in response to sand lance *Ammodytes spp.* behavior and distribution. Mar. Ecol. Prog. Ser. 395, 91–100.

Friedlaender, A., Tyson, R., Stimpert, A., Read, A. and Nowacek, D. (2013). Extreme diel variation in the feeding behavior of humpback whales along the western Antarctic Peninsula during autumn. Mar. Ecol. Prog. Ser. 494, 281–289.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1998). Bayesian Data Analysis. (ed. Chatfield, C. and Zidek, J. V) London, UK: Chapman & Hall.

Goldbogen, J. A., Calambokidis, J., Shadwick, R. E., Oleson, E. M., McDonald, M. A. and Hildebrand, J. A. (2006). Kinematics of foraging dives and lunge-feeding in fin whales. J. Exp. Biol. 209, 1231–1244.

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., Calambokidis, J., Friedlaender, A. S., Francis, J., DeRuiter, S. L., Stimpert, A. K., Falcone, E. and Southall, B. L. (2013). Underwater acrobatics

by the world's largest predator: 360° rolling manoeuvres by lunge-feeding blue whales. Biol. Lett. 9, 20120986.

Groves, P. D. (2013). Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems. 2nd ed. Boston, MA: Artech House.

Haddle, G. P. and Skudrzyk, E. J. (1969). The physics of flow noise. J. Acoust. Soc. Am. 46, 130–157.

Hastie, G. D., Gillespie, D. M., Gordon, J. C. D., Macaulay, J. D. J., Mcconnell, B. J. and Sparling, C. E. (2014). Tracking technologies for quantifying marine mammal interactions with tidal turbines: pitfalls and possibilities. In Marine Renewable Energy Technology and Environmental Interactions (ed. Shields, M. A. and Payne, A. I. L.), pp. 127–139. Dordrecht: Springer, The Netherlands.

Hazel, J. (2009). Evaluation of fast-acquisition GPS in stationary tests and fine-scale tracking of green turtles. J. Exp. Mar. Biol. Ecol. 374, 58–68.

Hoenner, X., Whiting, S. D., Hindell, M. A. and McMahon, C. R. (2012). Enhancing the use of Argos satellite data for home range and long distance migration studies of marine animals. PLoS One 7, e40713.

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

Jackson, J. E. (1991). A User's Guide to Principal Components. New York, NY: John Wiley & Sons.

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.

Johnson, M., Hickmott, L. S., Aguilar Soto, N. and Madsen, P. T. (2008a). Echolocation behaviour adapted to prey in foraging Blainville's beaked whale (*Mesoplodon densirostris*). Proc. R. Soc. B 275, 133–139.

Johnson, D. S., London, J. M., Lea, M. A. and Durban, J. W. (2008b). Continuous-time correlated random walk model for animal telemetry data. Ecology 89, 1208–1215.

Johnson, M., Aguilar de Soto, N. and Madsen, P. T. (2009). Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: a review. Mar. Ecol. Prog. Ser. 395, 55–73.

Jonsen, I. D., Flenming, J. M. and Myers, R. A. (2005). Robust state-space modeling of animal movement data. Ecology 86, 2874–2880.

Kuhn, C. E., Johnson, D. S., Ream, R. R. and Gelatt, T. S. (2009). Advances in the tracking of marine species: Using GPS locations to evaluate satellite track data and a continuous-time movement model. Mar. Ecol. Prog. Ser. 393, 97–109.

Kvadsheim, P., Lam, F.-P., Miller, P., Doksaeter, 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 Norwegain waters - 3S-2011 cruise report.

Kvadsheim, P., Lam, F.-P., Miller, P., Wensveen, P., Visser, F., Sivle, L., Kleivane, L., Curé, C., Ensor, P., van Ijsselmuide, S., et al. (2012). Behavioural responses of cetaceans to naval sonar signals in Norwegian waters – the 3S-2012 cruise report.

Laplanche, C. (2012). Bayesian three-dimensional reconstruction of toothed whale trajectories: passive acoustics assisted with visual and tagging measurements. J. Acoust. Soc. Am. 132, 3225–3233.

Laplanche, C., Marques, T. A. and Thomas, L. (2015). Tracking marine mammals in 3D using electronic tag data. Methods Ecol. Evol. 6, 987–996.

Leaper, R., Burt, L., Gillespie, D. and Macleod, K. (2010). Comparisons of measured and estimated distances and angles from sightings surveys. J. Cetacean Res. Manag. 11, 229–237.

Lerczak, J. A. and Hobbs, R. C. (1998). Calculating sighting distances from angular readings during shipboard, aerial, and shore-based marine mammal surveys. Mar. Mammal Sci. 14, 590–599.

Lisovski, S., Hewson, C. M., Klaassen, R. H. G., Korner-Nievergelt, F., Kristensen, M. W. and Hahn, S. (2012). Geolocation by light: Accuracy and precision affected by environmental factors. Methods Ecol. Evol. 3, 603–612.

Macfarlane, N. B. W., Howland, J. C., Jensen, F. H. and Tyack, P. L. (2015). A 3D stereo camera system for precisely positioning animals in space and time. Behav. Ecol. Sociobiol. 69, 685–693.

Marques, T. A. (2004). Predicting and correcting bias caused by error measurement in line transect sampling using multiplicative error models. Biometrics 60, 757–763.

Mate, B. (2012). The implementation of acoustic dosimeters with recoverable month-long GPS/TDR tags to interpret controlled exposure experiments for large whales. Adv. Exp. Med. Biol. 730, 203–205.

Matsumura, M., Watanabe, Y. Y., Robinson, P. W., Miller, P. J. O., Costa, D. P. and Miyazaki, N. (2011). Underwater and surface behavior of homing juvenile northern elephant seals. J. Exp. Biol. 214, 629–636.

McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J. and Morales, J. M. (2012). A general discrete-time modeling framework for animal movement using multistate random walks. Ecol. Monogr. 82, 335–349.

McClintock, B. T., London, J. M., Cameron, M. F. and Boveng, P. L. (2014a). Modelling animal movement using the Argos satellite telemetry location error ellipse. Methods Ecol. Evol. 6, 266–277.

McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M. and Morales, J. M. (2014b). When to be discrete: the importance of time formulation in understanding animal movement. Mov. Ecol. 2, 21.

McConnell, B. J., Chambers, C. and Fedak, M. A. (1992). Foraging ecology of southern elephant seals in relation to the bathymetry and productivity of the Southern Ocean. Antarct. Sci. 4, 393–398.

McConnell, B., Beaton, R., Bryant, E., Hunter, C., Lovell, P. and Hall, A. (2004). Phoning home-a new GSM mobile phone telemetry system to collect mark-recapture data. Mar. Mammal Sci. 20, 274–283.

Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P. and Matsumoto, H. (2007). An overview of fixed passive acoustic observation methods for cetaceans. Oceanography 20, 37–45.

Miller, P. J., Johnson, M. P., Tyack, P. L. and Terray, E. A. (2004). Swimming gaits, passive drag and buoyancy of diving sperm whales *Physeter macrocephalus*. J. Exp. Biol. 207, 1953–1967.

Miller, P. J., Johnson, M. P., Madsen, P. T., Biassoni, N., Quero, M. and Tyack, P. L. (2009). Using at-sea experiments to study the effects of airguns on the foraging

behavior of sperm whales in the Gulf of Mexico. Deep Sea Res. I 56, 1168–1181.

Miller, P. J. O., Kvadsheim, P. H., Lam, F.-P. A., Wensveen, P. J., Antunes, R., Alves, A. C., Visser, F., Kleivane, L., Tyack, P. L. 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.

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

Miller, P. J. O., Kvadsheim, P. H., Lam, F. P. A., Tyack, P. L., Curé, C., DeRuiter, S. L., Kleivane, L., Sivle, L. D., van IJsselmuide, S. P., Visser, F., et al. (2015). First indications that northern bottlenose whales are sensitive to behavioural disturbance from anthropogenic noise. R. Soc. Open Sci. 2, 140484.

Mitani, Y., Andrews, R. D., Sato, K., Kato, A., Naito, Y. and Costa, D. P. (2010). Three-dimensional resting behaviour of northern elephant seals: drifting like a falling leaf. Biol. Lett. 6, 163–166.

Moore, M. J., Miller, C. A., Morss, M. S., Arthur, R., Lange, W. A., Prada, K. G., Marx, M. K. and Frey, E. A. (2001). Ultrasonic measurement of blubber thickness in right whales. J. Cetacean Res. Manag. Spec Iss 2, 301–309.

Muramoto, H., Ogawa, M., Suzuki, M. and Naito, Y. (2004). Little Leonardo digital data logger: its past, present and future roll in bio-logging science. Mem. Natl. Inst. Polar Res., Spec Issue 58, 196–202.

Narazaki, T., Sato, K., Abernathy, K. J., Marshall, G. J. and Miyazaki, N. (2009). Sea turtles compensate deflection of heading at the sea surface during directional travel. J. Exp. Biol. 212, 4019–4026.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. and Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. Proc. Natl. Acad. Sci. U. S. A. 105, 19052–19059.

Nielsen, A., Bigelow, K. A., Musyl, M. K. and Sibert, J. R. (2006). Improving light-based geolocation by including sea surface temperature. Fish. Oceanogr. 15, 314–325.

Parks, S. E., Cusano, D. A., Stimpert, A. K., Weinrich, M. T., Friedlaender, A. S. and Wiley, D. N. (2014). Evidence for acoustic communication among bottom foraging humpback whales. Sci. Rep. 4, 7508.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O. and Matthiopoulos, J. (2008). State-space models of individual animal movement. Trends Ecol. Evol. 23, 87–94.

Patterson, T. A., McConnell, B. J., Fedak, M. A., Bravington, M. V and Hindell, M. A. (2010). Using GPS data to evaluate the accuracy of state-space methods for correction of Argos satellite telemetry error. Ecology 91, 273–285.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, pp. 1–10. Vienna, Austria.

R Core Team (2013). R: A language and environment for statistical computing.

Samarra, F. I. P. and Miller, P. J. O. (2015). Prey-induced behavioural plasticity of herring-eating killer whales. Mar. Biol. 162, 809–821.

Schmidt, V., Weber, T. C., Wiley, D. N. and Johnson, M. P. (2010). Underwater tracking of humpback whales (*Megaptera novaeangliae*) with high-frequency pingers and acoustic recording tags. IEEE J. Ocean. Eng. 35, 821–836.

Schofield, G., Bishop, C. M., MacLean, G., Brown, P., Baker, M., Katselidis, K. A., Dimopoulos, P., Pantis, J. D. and Hays, G. C. (2007). Novel GPS tracking of sea turtles as a tool for conservation management. J. Exp. Mar. Bio. Ecol. 347, 58–68.

Shimada, T., Jones, R., Limpus, C. and Hamann, M. (2012). Improving data retention and home range estimates by data-driven screening. Mar. Ecol. Prog. Ser. 457, 171–180.

Shiomi, K., Sato, K., Mitamura, H., Arai, N., Naito, Y. and Ponganis, P. J. (2008). Effect of ocean current on the dead-reckoning estimation of 3-D dive paths of emperor penguins. Aquat. Biol. 3, 265–270.

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.

Sirtrack (2012). Fastloc 2 User Guide.

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.

Stanistreet, J. E., Risch, D. and Van Parijs, S. M. (2013). Passive acoustic tracking of singing humpback whales (*Megaptera novaeangliae*) on a northwest Atlantic feeding ground. PLoS One 8, e61263.

Stimpert, A. K., Wiley, D. N., Au, W. W. L., Johnson, M. P. and Arsenault, R. (2007). "Megapclicks": acoustic click trains and buzzes produced during night-time foraging of humpback whales (*Megaptera novaeangliae*). Biol. Lett. 3, 467–470.

Sumner, M. D., Wotherspoon, S. J. and Hindell, M. A. (2009). Bayesian estimation of animal movement from archival and satellite tags. PLoS One 4, e7324.

Tomkiewicz, S. M., Fuller, M. R., Kie, J. G. and Bates, K. K. (2010). Global positioning system and associated technologies in animal behaviour and ecological research. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365, 2163–2176.

Tyack, P., Gordon, J. and Thompson, D. (2003). Controlled-exposure experiments to determine the effects of noise on marine mammals. Mar. Technol. Soc. J. 37, 39–51.

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

Tyson, R. B., Friedlaender, A. S., Ware, C., Stimpert, A. K. and Nowacek, D. P. (2012). Synchronous mother and calf foraging behaviour in humpback whales *Megaptera novaeangliae*: insights from multi-sensor suction cup tags. Mar. Ecol. Prog. Ser. 457, 209–220.

Tremblay, Y., Shaffer, S. A., Fowler, S. L., Kuhn, C. E., McDonald, B. I., Weise, M. J., Bost, C.-A., Weimerskirch, H., Crocker, D. E., Goebel, M. E., et al. (2006).

Interpolation of animal tracking data in a fluid environment. J. Exp. Biol. 209, 128–140.

Venables, W. N. and Ripley, B. D. (2002). Modern Applied Statistics with S. 4th ed. New York, NY: Springer.

Vincent, C., Mcconnell, B. J., Ridoux, V. and Fedak, M. A. (2002). Assessment of ARGOS location accuracy from satellite tags deployed on captive gray seals. Mar. Mammal Sci. 18, 156–166.

Von Benda-Beckmann, A. M., Beerens, S. P. and van IJsselmuide, S. P. (2013). Effect of towed array stability on instantaneous localization of marine mammals. J. Acoust. Soc. Am. 134, 2409–2417.

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.

Ward, J., Morrissey, R., Moretti, D., DiMarzio, N., Jarvis, S., Johnson, M., Tyack, P. and White, C. (2008). Passive acoustic detection and localization of *Mesoplodon densirostris* (Blainville's beaked whale) vocalizations using bottom-mounted hydrophones in conjuction with a Digital Tag (DTag) recording. Can. Acous. 36, 60–66.

Ware, C., Friedlaender, A. S. and Nowacek, D. P. (2011). Shallow and deep lunge feeding of humpback whales in fjords of the West Antarctic Peninsula. Mar. Mammal Sci. 27, 587–605.

Wartzok, D., Sayegh, S., Stone, H., Barchak, J. and Barnes, W. (1992). Acoustic tracking system for monitoring under-ice movements of polar seals. J. Acoust. Soc. Am. 92, 682–687.

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. (2015). 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.

Wiley, D., Ware, C., Bocconcelli, A., Cholewiak, D., Friedlaender, A., Thompson, M. and Weinrich, M. (2011). Underwater components of humpback whale bubble-net feeding behaviour. Behaviour 148, 575–602.

Williams, R., Leaper, R., Zerbini, A. N. and Hammond, P. S. (2007). Methods for investigating measurement error in cetacean line-transect surveys. J. Mar. Biol. Ass. UK 87, 313–320.

Wilson, R. P. and Wilson, M. P. (1988). Dead reckoning: a new technique for determining penguin movements at sea. Meeresforschung 32, 155–158.

Wilson, R. P., Liebsch, N., Davies, I. M., Quintana, F., Weimerskirch, H., Storch, S., Lucke, K., Siebert, U., Zankl, S., Müller, G., et al. (2007). All at sea with animal tracks; methodological and analytical solutions for the resolution of movement. Deep Sea Res II 54, 193–210.

Wilson, R. P., Shepard, E. L. C. and Liebsch, N. (2008). Prying into the intimate details of animal lives: use of a daily diary on animals. Endanger. Species Res. 4, 123–137.

Winship, A. J., Jorgensen, S. J., Shaffer, S. A., Jonsen, I. D., Robinson, P. W., Costa, D. P. and Block, B. A. (2012). State-space framework for estimating measurement error from double-tagging telemetry experiments. Methods Ecol. Evol. 3, 291–302.

Witt, M. J., Åkesson, S., Broderick, A. C., Coyne, M. S., Ellick, J., Formia, A., Hays, G. C., Luschi, P., Stroud, S. and Godley, B. J. (2010). Assessing accuracy and utility of satellite-tracking data using Argos-linked Fastloc-GPS. Anim. Behav. 80, 571–581.

Witteveen, B. H., Foy, R. J., Wynne, K. M. and Tremblay, Y. (2008). Investigation of foraging habits and prey selection by humpback whales (*Megaptera novaeangliae*) using acoustic tags and concurrent fish surveys. Mar. Mammal Sci. 24, 516–534.

Zidek, J., Liu, Y. and Trites, A. (2014). Characterizing uncertainty in the bio-logged paths of sea mammals. In *Abstracts of the 5th International Bio-logging Science Symposium, 22-27 September*, p. 160. Strasbourg, France.

Zimmer, W. M., Johnson, M. P., Madsen, P. T. and Tyack, P. L. (2005). Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*). J. Acoust. Soc. Am. 117, 3919–3927.

TABLES & FIGURES

Table 2.1. Summary of the data sets. For each humpback whale are given the IDs of the DTAG and Fastloc-GPS loggers, geographical coordinates of the initial observed position, track duration, number of position fixes obtained by visual observation and Fastloc-GPS, and computational runtime of the model. The DTAG ID contains information about the species, year, Julian day and tag-of-day; for example, 'mn11_157a' refers to the first tag ('a') deployed on a humpback whale (*Megaptera novaeangliae*) on day 157 of 2011.

Whale	**DTAG ID**	**FGPS ID**	**Initial position**		**Track duration**	**Position fixes**		**Runtime**
			Lat. °N	**Long.** °E	h	**Visual** #	**FGPS** #	h
1	mn11_157a	29420	75.141	14.603	14.7	105	451	22.5
2	mn11_158a	29409	74.832	16.715	7.6	70	20	0.8
3	mn11_160a	29409	74.651	15.236	13.0	116	0	1.0
4	mn11_165e	29409	78.074	11.824	11.3	123	205	7.6
5	mn11_176b	-	77.563	12.537	2.9	44	-	0.2
6	mn12_161a	29420, 29510	77.556	11.277	10.9	186	0	2.6
7	mn12_164a	29409	77.798	10.073	7.7	122	391	17.1
8	mn12_164b	29409	77.824	9.793	3.8	68	100	2.3
9	mn12_170b	29409	77.512	11.633	8.3	87	249	21.6
10	mn12_171b	29409, 29510	79.032	10.612	7.8	127	646	78.4
11	mn12_178a	29420	74.867	17.767	7.6	50	159	4.2
12	mn12_179a	29420	74.051	20.675	8.5	75	202	6.0
13	mn12_180b	29409, 29420	73.993	20.398	7.6	115	730	54.7

Table 2.2. Prior probability distributions for all parameters estimated. Uniform priors were assumed for $\boldsymbol{\sigma}$ and $\boldsymbol{v}_1^{cor}$. Prior distributions for the initial unobserved whale position $\boldsymbol{x}_1$ reflected the prior knowledge about the accuracy of the initial observed position (at coordinates *x*=0, *y*=0). These priors therefore depended upon whether the position was observed (1) using Fastloc-GPS or (2) visually. Values for the priors on $\boldsymbol{\sigma}$ and $\boldsymbol{v}_1^{cor}$ are in metres per second.

Parameter	**Description**	**Prior**
σ_x	Process error standard deviation, *x*-dimension	$Unif(0, 0.1)$
σ_y	Process error standard deviation, *y*-dimension	$Unif(0, 0.1)$
$v_{x,1}^{cor}$	Initial velocity correction, *x*-dimension	$Unif(-1, 1)$
$v_{y,1}^{cor}$	Initial velocity correction, *y*-dimension	$Unif(-1, 1)$
$x_{x,1}$	Initial whale position, *x*-dimension (1)	$t(0, \sigma_{x,q}^F, \upsilon_{x,q}^F)$
	Initial whale position, *x*-dimension (2)	$N(0, R_1\sigma_m^r/100)$
$x_{y,1}$	Initial whale position, *y*-dimension (1)	$t(0, \sigma_{y,q}^F, \upsilon_{y,q}^F)$
	Initial whale position, *y*-dimension (2)	$N(0, R_1\sigma_m^r/100)$

Table 2.3. Fastloc-GPS accuracy test results. Scaled *t* distributions were fitted to the positional errors measured during the Fastloc-GPS accuracy tests. Maximum likelihood estimates and standard error (SEs) are provided for location $\boldsymbol{\mu}^F$, scale $\boldsymbol{\sigma}^F$, and shape $\boldsymbol{\upsilon}^F$ for each spatial dimension and number of satellites used for the position calculation (#satellite bin).

Parameter	**Number of satellites**	***X*-dimension**		***Y*-dimension**	
		Estimate	*SE*	*Estimate*	*SE*
$\boldsymbol{\mu}^F$ (m)	4	0.57	0.56	−1.06	0.76
	5	0.21	0.38	−0.41	0.49
	6	0.02	0.22	0.17	0.28
	7	0.01	0.16	0.10	0.20
	8	−0.01	0.14	0.39	0.17
	>8	0.08	0.09	−0.01	0.11
$\boldsymbol{\sigma}^F$ (m)	4	24.51	0.68	34.07	0.03
	5	19.11	0.42	25.37	0.04
	6	13.10	0.23	17.12	0.10
	7	10.69	0.16	14.23	0.19
	8	9.28	0.14	11.56	0.42
	>8	7.77	0.10	9.35	0.65
$\boldsymbol{\upsilon}^F$ (-)	4	0.93	0.90	1.08	0.03
	5	1.44	0.55	1.64	0.05
	6	2.53	0.29	2.73	0.11
	7	3.91	0.21	5.32	0.34
	8	5.83	0.18	6.86	0.58
	>8	8.17	0.12	7.72	0.61

Table 2.4. Mean, standard deviation (SD) and 95% credibility interval (CI) of the marginal posterior probability distributions for parameters σ_x, σ_y, $v_{x,1}^{cor}$, $v_{y,1}^{cor}$, $x_{x,1}$, $x_{y,1}$, and the values of convergence statistic $\hat{R}$.

Parameter	**Statistic**	**Whale**					
		1	**2**	**3**	**4**	**5**	**6**
sigma_x (m/s)	mean	0.006	0.010	0.009	0.007	0.008	0.006
	SD	0.001	0.002	0.001	0.001	0.003	0.001
	95% CI	0.004	0.007	0.007	0.005	0.004	0.004
		0.008	0.014	0.012	0.009	0.015	0.007
	Rhat	1.02	1.02	1.02	1.01	1.00	1.00
sigma_y (m/s)	mean	0.005	0.009	0.010	0.007	0.015	0.007
	SD	0.001	0.001	0.001	0.001	0.004	0.001
	95% CI	0.004	0.006	0.008	0.005	0.009	0.005
		0.006	0.011	0.012	0.009	0.023	0.009
	Rhat	1.00	1.00	1.05	1.00	1.00	1.01
vcor_x1 (m/s)	mean	-0.26	0.54	-0.34	0.15	0.49	-0.15
	SD	0.08	0.18	0.07	0.10	0.10	0.07
	95% CI	-0.42	0.14	-0.47	-0.03	0.28	-0.29
		-0.12	0.86	-0.18	0.35	0.68	-0.01
	Rhat	1.00	1.00	1.00	1.00	1.01	1.00
vcor_y1 (m/s)	mean	0.07	0.19	0.57	0.13	-0.77	-0.30
	SD	0.11	0.17	0.08	0.07	0.12	0.10
	95% CI	-0.41	0.22	-0.40	0.12	0.30	-0.22
		-0.13	0.67	-0.10	0.37	0.70	0.00
	Rhat	1.00	1.00	1.00	1.00	1.00	1.00
x_x1 (m)	mean	6.1	3.2	0.8	-5.2	16.4	-0.1
	SD	12.6	14.8	9.3	9.2	78.2	5.9
	95% CI	-16.3	-24.9	-15.7	-26.4	-118.4	-12.0
		35.2	34.9	20.6	10.5	188.7	11.6
	Rhat	1.00	1.00	1.00	1.00	1.01	1.00
x_y1 (m)	mean	-63.0	6.8	2.1	0.5	-13.3	-0.8
	SD	26.3	21.1	9.0	9.4	62.5	5.7
	95% CI	-47.0	-2.8	-98.9	-38.0	1114.9	-135.2
		-7.0	92.4	-37.9	29.8	1313.4	-77.3
	Rhat	1.00	1.00	1.00	1.00	1.02	1.00

Table 2.4. Continued

Track ID						
7	**8**	**9**	**10**	**11**	**12**	**13**
0.007	0.006	0.004	0.014	0.014	0.005	0.010
0.001	0.001	0.000	0.001	0.001	0.001	0.001
0.006	0.004	0.003	0.012	0.012	0.004	0.008
0.009	0.008	0.004	0.015	0.017	0.007	0.011
1.00	1.00	1.00	1.01	1.00	1.00	1.00
0.006	0.007	0.003	0.012	0.012	0.012	0.009
0.001	0.001	0.000	0.001	0.001	0.002	0.001
0.005	0.005	0.003	0.010	0.010	0.009	0.007
0.008	0.011	0.004	0.013	0.015	0.015	0.010
1.00	1.00	1.00	1.02	1.00	1.01	1.00
0.13	-0.18	0.13	-0.25	0.09	0.06	0.18
0.08	0.09	0.05	0.18	0.04	0.05	0.07
-0.02	-0.35	0.03	-0.60	0.02	-0.03	0.05
0.29	-0.01	0.24	0.11	0.17	0.16	0.31
1.01	1.00	1.00	1.00	1.00	1.00	1.01
-0.27	0.19	0.22	-0.19	0.11	-0.04	0.07
0.09	0.11	0.06	0.17	0.05	0.09	0.07
0.00	-0.25	0.03	-0.62	-0.06	0.03	0.05
0.27	-0.09	0.23	0.03	0.25	0.20	0.29
1.00	1.00	1.00	1.01	1.00	1.00	1.00
-1.5	-2.8	-6.4	-15.7	0.6	-3.1	8.0
6.2	8.5	7.0	20.7	15.4	9.8	8.5
-13.7	-21.8	-21.2	-72.4	-33.3	-24.6	-7.9
11.0	12.9	6.8	16.0	29.7	15.1	25.5
1.01	1.00	1.00	1.02	1.00	1.00	1.01
-12.9	-4.0	20.5	31.7	-2.1	-2.0	-3.2
11.1	10.0	9.4	31.9	18.7	13.0	10.0
-13.2	-12.2	-51.8	-141.6	41.6	-93.8	-27.1
8.7	21.3	-26.9	-61.6	75.0	-47.6	-9.0
1.00	1.00	1.00	1.00	1.00	1.00	1.00

Figure 2.1. Boxplots of the Fastloc-GPS positional errors for the three data loggers (29409, 29402, and 29520, from top to bottom), four calibration test sites (A: 56.33°N, 2.78°W; B: 69.68°N, 18.99°E; C: 78.24°N, 15.54°E; D: 64.92°N, 23.25°W), and six #satellite bins (4, 5, 6, 7, 8, and >8). The sample size for each subset is indicated on the right vertical axis. Outlier data points were omitted to improve readability.

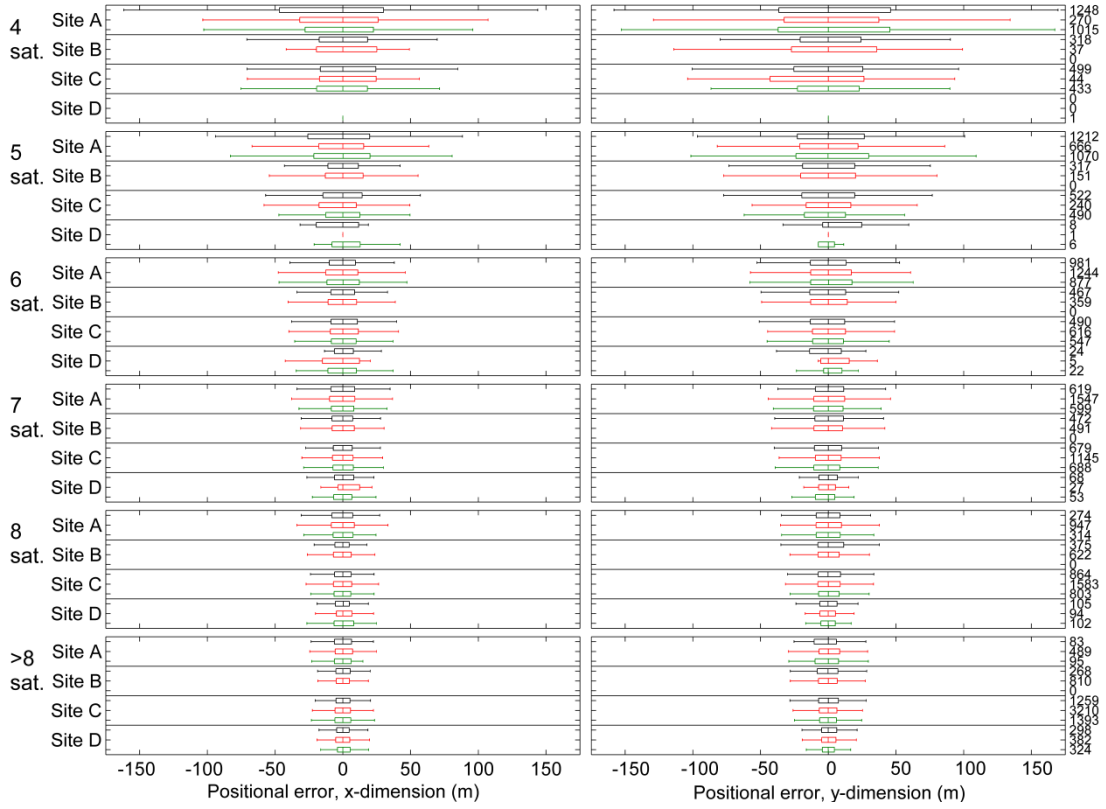


Figure 2.2. Error distributions from Fastloc-GPS accuracy tests. Scaled histograms (grey bins) of the Fastloc-GPS positional errors and the corresponding pdfs (black lines) of the scaled *t* distributions are shown as functions of spatial dimensions *x* and *y* and the number of satellites used in the position calculation. All graphs are truncated at ±130 m for clarity, although positional errors of several kilometres were occasionally observed.

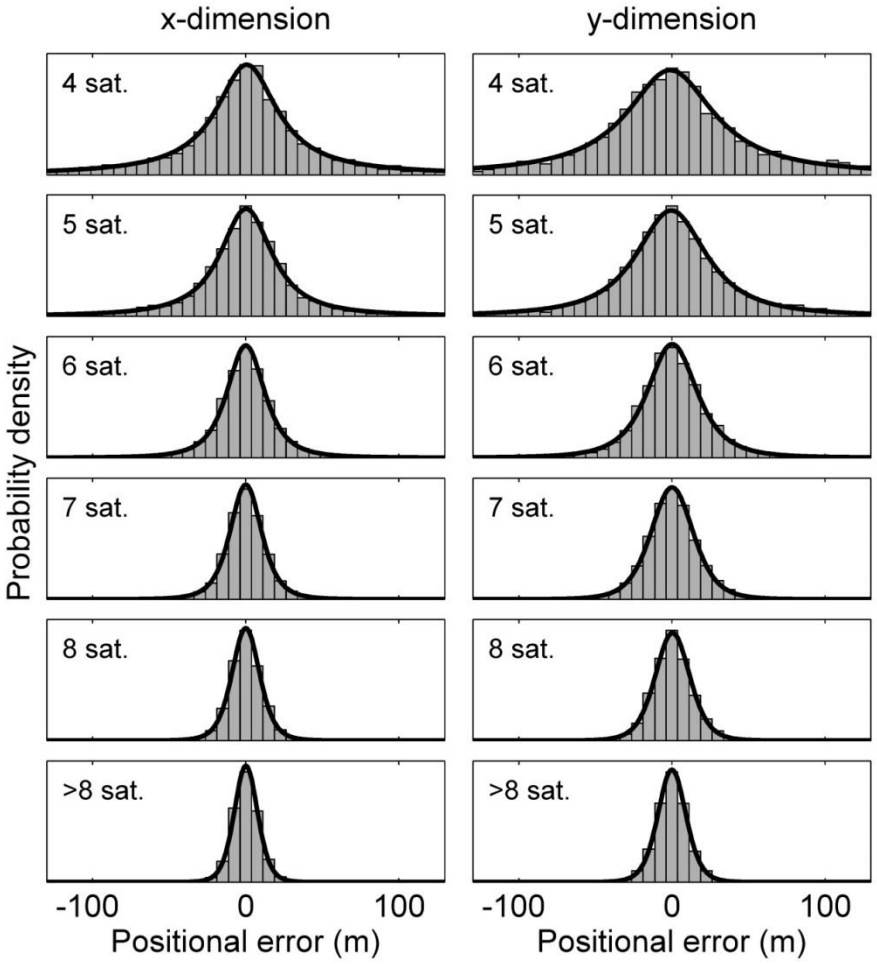


Figure 2.3. Error distributions from visual accuracy tests. Distributions of (A) angular errors in the bearing from the observer to the whale, and (B) errors in range expressed as a percentage of true range. Grey bins represent scaled histograms of the observation errors and black lines represent the pdfs of the fitted distributions (wrapped Cauchy for bearing; Normal for range). The scatterplots in the right panels illustrate: (C) the range estimated by the observers during tests as function of the true range derived from GPS positions, and (D) the range percent errors vs. true range, with the fitted linear regression line indicating little tendency for under- or overestimation.

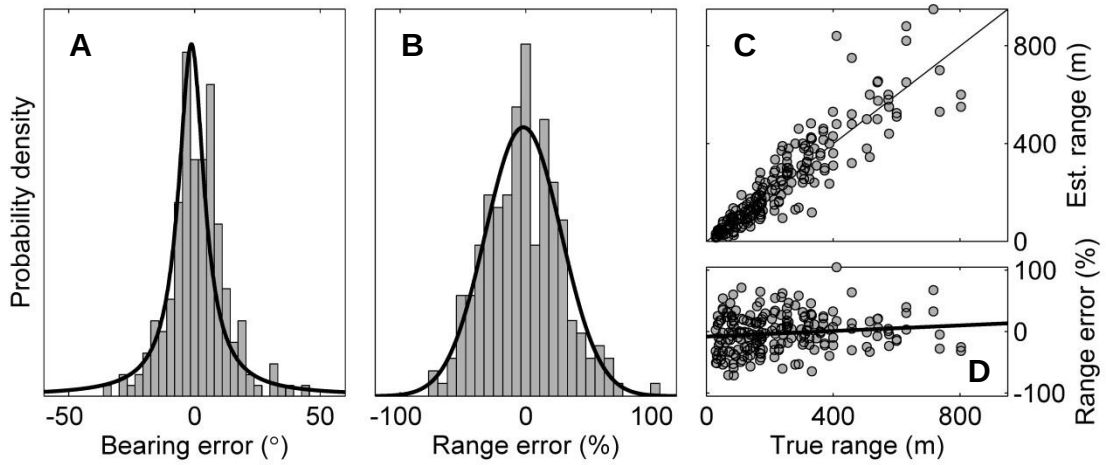


Figure 2.4. Example of a reconstructed track. Shown are (A) the full, most probable track (i.e. the posterior means of $\boldsymbol{x}$) and position fixes recorded for humpback whale 11 (i.e. mn12_178a) and (B) a detailed view of sections of the track. Visual position fixes were derived from ranges that were estimated by eye or measured using a laser range finder (LRF). Information only shown in (B): the GPS positions of the observation boat, 10% of the computed whale track realisations, and the most probable whale positions at the times of the fixes (t_j) with their 95% confidence ellipses (Jackson, 1991). Movement parameters of the track are shown in the panels on the right: (from top to bottom) the whale's body pitch and heading angles measured in the Earth frame, the whale's speed-through-water derived from flow noise, the uncorrected velocity of the whale, the posterior mean velocity correction with 95% credibility intervals (CIs), and the depth of the whale (z-axis coordinate of its position).

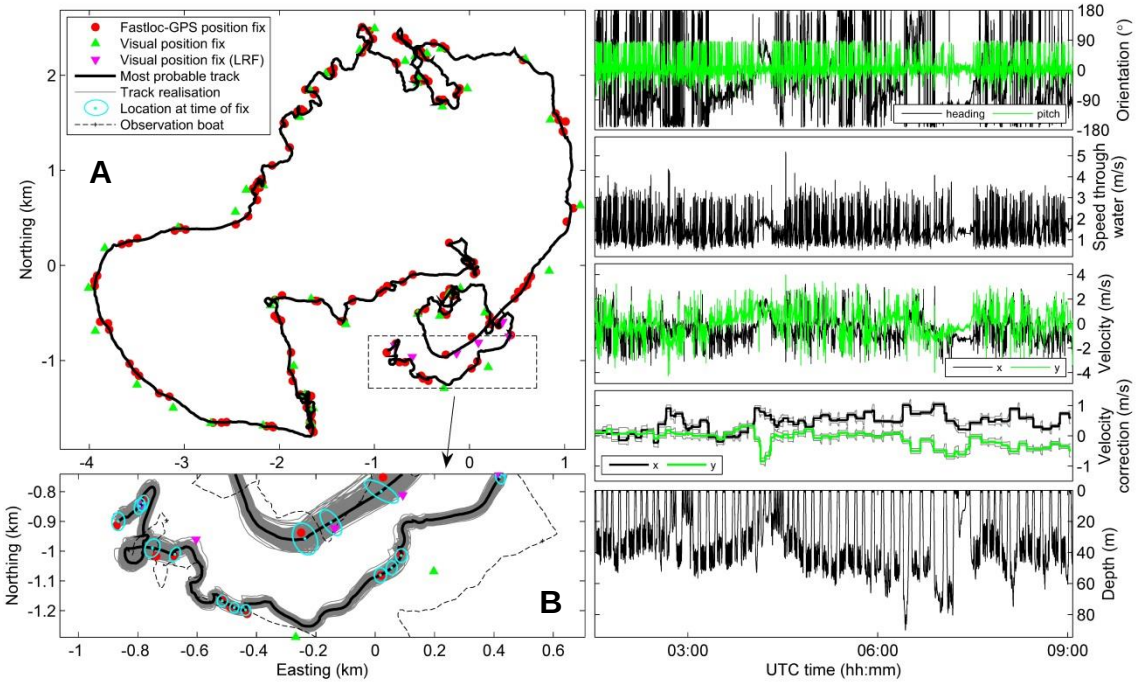


Figure 2.5. Marginal posterior probability distributions for all whale tracks.

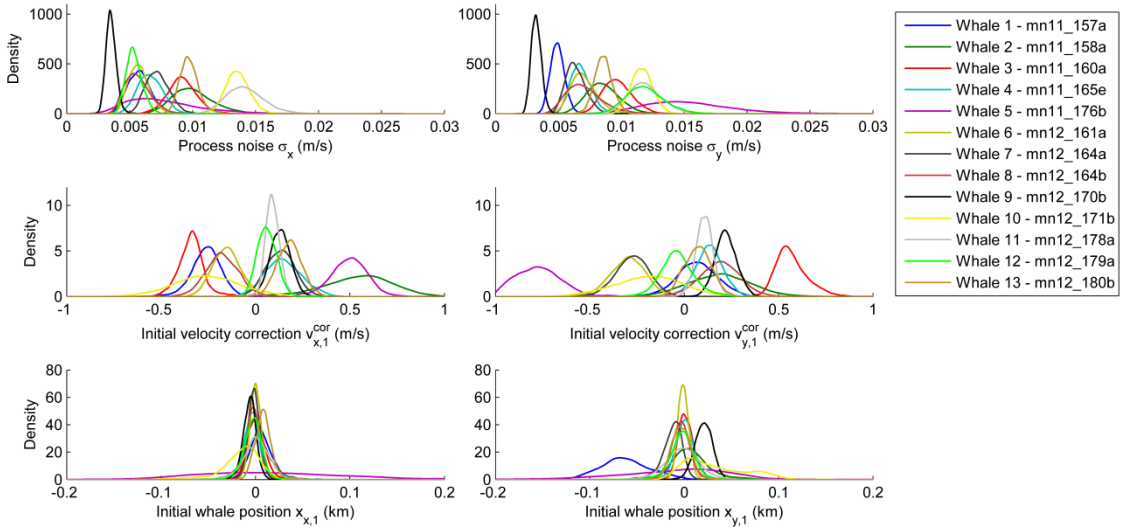


Figure 2.6. Results of the cross-validation analysis. Cross-validation errors (mean ± 2 s.e.m.) are shown as function of the number of satellites of the validation set (i.e. the out-of-set Fastloc-GPS data) for analyses where (A) single positions were omitted (10% of data) and (B) series of five consecutive positions were omitted (50% of data). Positional cross-validation errors were calculated for five different track types: 1) a track with linear interpolation between visual position fixes (♦), 2) a track with linear interpolation between Fastloc-GPS position fixes (×), 'forced-point' dead-reckoning tracks initially calculated with 3) constant speed (■) or 4) speed from flow noise (▼), and 5) the mean posterior track of the Bayesian state-space model (●). One-dimensional positional errors for Fastloc-GPS derived from the large data set collected during dry tests (▲) are shown for comparison (see also Fig. 2.7). Symbol horizontal positions have been offset for clarity.

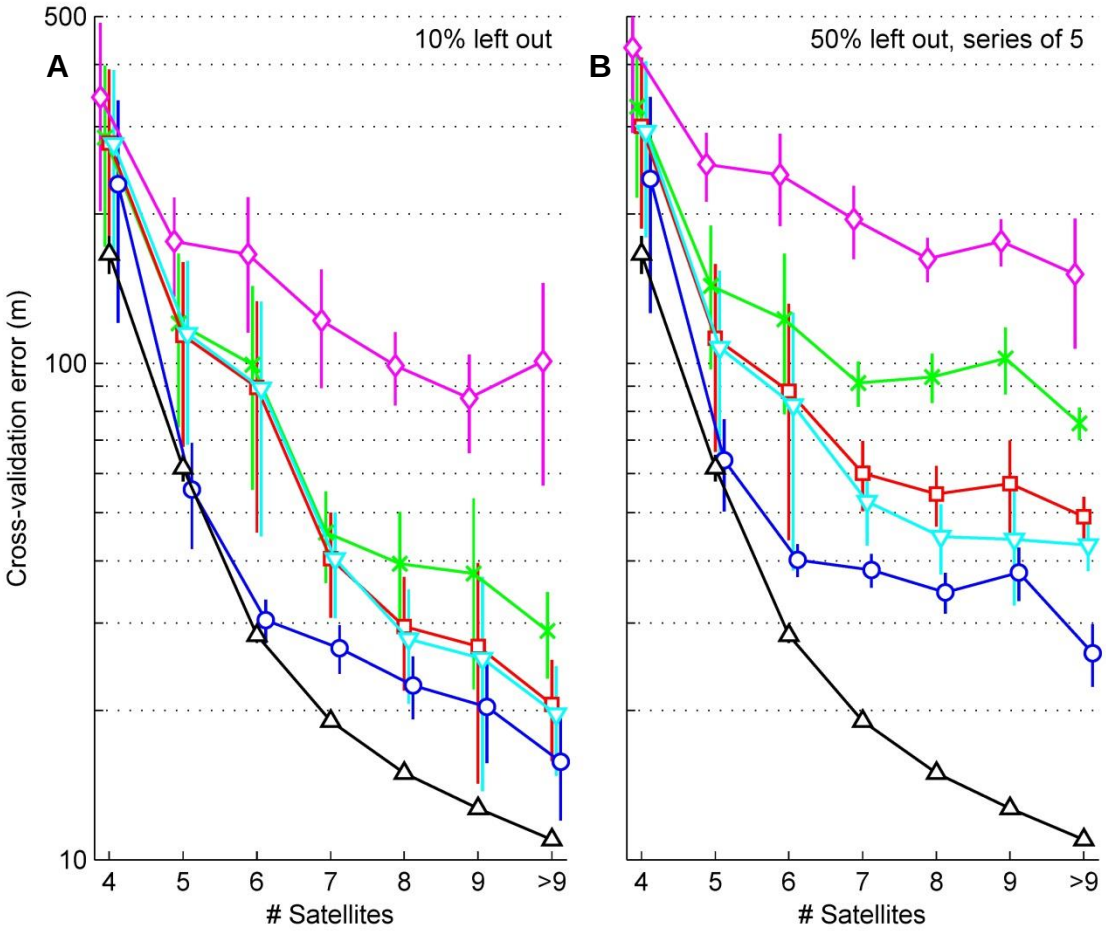
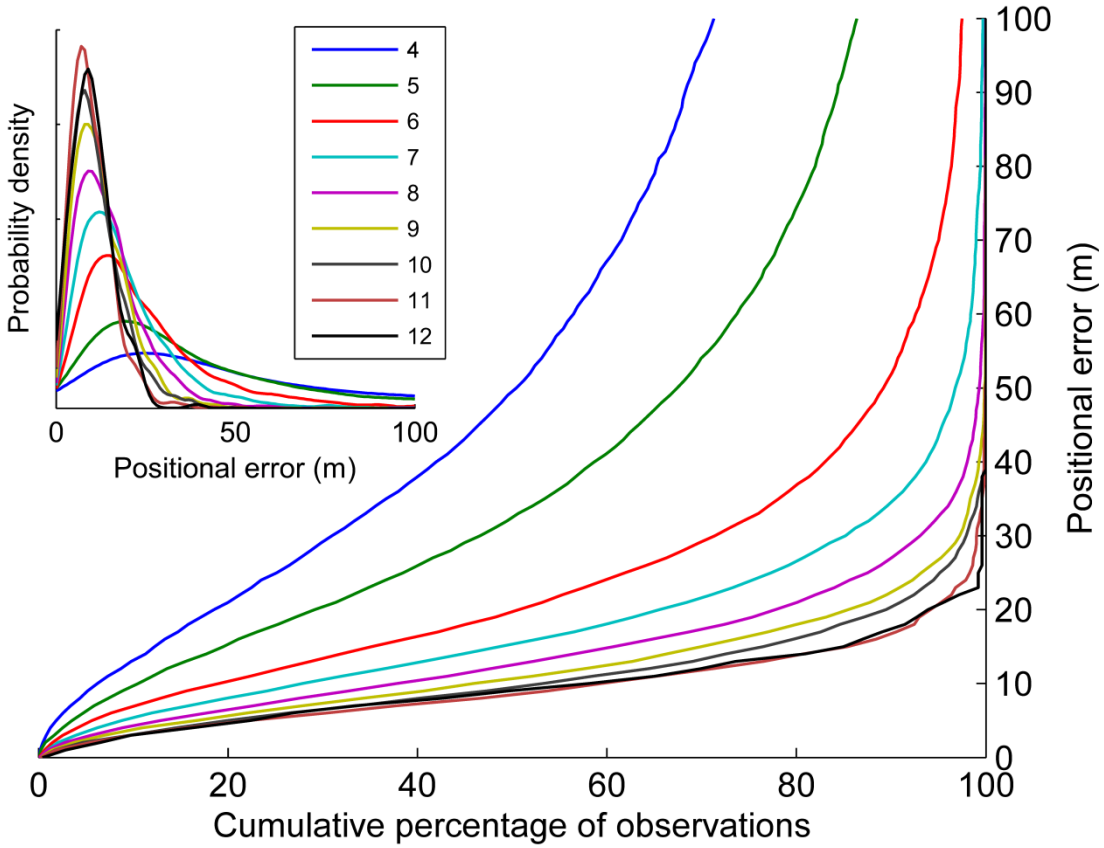


Figure 2.7. One-dimensional Fastloc-GPS errors. Positional errors during calibrations were represented as radial distances from the median and plotted against the cumulative percentage of positions for comparison with other studies. Each line represents a subset of data based upon the number of satellites (4 to 12) used for the position calculation. The insert shows the pdfs for the 9 satellite coverage categories. The graphs were truncated at 100 m for clarity.



Chapter 3

Dose-response functions for avoidance onset and feeding cessation by humpback whales in response to naval sonar

SUMMARY

There is concern that anthropogenic noise may significantly impact the habitats and welfare of marine mammals. Naval active sonar has received particular attention due to several unusual mass strandings of primarily beaked whales, and because the high sound levels that these systems can produce may disrupt important behaviours. Naval sonars may also affect humpback whales (*Megaptera novaeangliae*), which have a global distribution and vulnerable conservation status, because of the overlap of sonar signals with their expected hearing range. In this chapter, dose-response functions for behavioural responses to sonar were derived from response thresholds identified during 18 controlled sonar exposure sessions with 10 independent groups of 1 or 2 whales. To obtain thresholds for the onset of horizontal avoidance and cessation of feeding, whales were instrumented with DTAGs and then exposed to repeated sonar signals (1.3-2 kHz) during dose-escalation experiments. Dose-response functions generated in a Bayesian analysis predicted that 50% of the humpbacks would initiate horizontal avoidance after receiving a sound pressure level (SPL) of 166 dB re 1 μPa (95% credibility interval (CI): 151-176), cumulative sound exposure level (SEL_{cum}) of 168 dB re 1 μPa^2 s (95% CI: 148-179), signal-to-noise ratio (S/N) of 71 dB (95% CI: 56-86) and source-whale distance of 578 m (95% CI: 427-842). Similar results were obtained for cessation of feeding, but the two response types were linked only in 25% of sessions. Although there were indications that session order and the animal's feeding state affected the response thresholds, small sample size and high variability in thresholds did not allow firm conclusions about the effects of these contextual factors. The dose-response function for SPL was generally consistent with humpback whale responsiveness to noise reported in literature. Comparison with published dose-response functions for other marine

mammals indicated that humpback whales are relatively unresponsive; however, some whales were predicted to respond at relatively low received levels.

INTRODUCTION

Active sonar is used by navies to detect and track quiet submarines, torpedoes and other objects under water. Most naval sonars produce high source levels compared to other sound sources and transmit repeated sound pulses in frequency bands that fall within the hearing ranges of most marine mammals (Ainslie, 2010; Mooney et al., 2012). Naval sonar has received particular attention due to several atypical mass strandings of predominantly, but not exclusively, deep-diving beaked whales, in connection with naval operations (Balcomb and Claridge, 2001; D'Amico et al., 2009; Fernández et al., 2005). While the exact causes of these mass strandings remain unknown, the proposed mechanisms include behavioural avoidance responses leading to stranding directly and/or maladaptive dive responses leading to tissue damage (Cox et al., 2006). Sound from naval sonar and other anthropogenic sources may also increase stranding rates of small cetaceans by impeding on their ability to detect and avoid fishing nets (Wright et al., 2013).

The fates of populations rather than individuals are eventually of most concern for conservation of marine mammal species, and the potential of sublethal effects to translate into population consequences has received increased recognition (National Research Council, 2005). Behavioural and physiological disturbance might increase the risk of population-level effects (e.g. decreased carrying capacity) if it affects an individual's health, survival, or ability to reproduce. As such, disturbance such as noise exposure may become biologically significant if it induces changes in energy intake or expenditure, social disruption, or avoidance of an area that is important for life functions such as feeding, breeding, or resting (New et al., 2013a; Pirotta et al., 2015; Schick et al., 2013). While direct and immediate effects (e.g. the separation of a dependent calf from its mother) may impact individual fitness, effects of sonar exposure that persist over longer temporal and larger spatial scales (Kuningas et al., 2013; McCarthy et al., 2011) are more likely to have the greatest impact. However, long-term behavioural effects are difficult to observe in free-ranging marine mammals and these

responses could be modified by learning (e.g. habituation, sensitisation, tolerance; Bejder et al., 2006, 2009), masking (Clark et al., 2009), hearing loss (Finneran, 2015) and stress responses (Atkinson et al., 2015).

One of the first steps in the causal chain of events from sonar exposure to possible population-level impacts or strandings is the relationship between the dose of the sonar stimulus and the behavioural response, which is quantified by a probabilistic dose-response function. Dose-response functions have been reported in recent scientific literature for several species. These studies had different experimental protocols, signal types, response indicators and levels of control over the stimulus and study animals. Experimental studies conducted in laboratory conditions showed that the location and shape of dose-response functions can depend on the frequency of the sonar (Kastelein et al., 2012) and age distribution (Houser et al., 2013b) and exposure history of the animals (Houser et al., 2013a). A dose-response function for disruption of foraging behaviour in Blainville's beaked whales was derived from passive acoustic monitoring data collected during actual multi-ship sonar operations (Moretti et al., 2014). In that study, cessation of clicking responses occurred at slightly higher levels than were observed during experimental exposures with tagged Blainville's and other beaked whales (DeRuiter et al., 2013; Miller et al., 2015a; Tyack et al., 2011), but few whales were exposed to relatively low received levels of sonar. Experimental studies conducted in free-ranging conditions with killer whales and long-finned pilot whales identified high levels of within- and between-variation in avoidance response thresholds and suggested lower avoidance thresholds for killer whales (Antunes et al., 2014; Miller et al., 2014).

Specific methodologies have been developed to collect data that are suitable for the construction of dose-response functions for behavioural responses. To assign the correct dose to a response, the dose is slowly increased throughout the experiment (dose-escalation) or a different dose is tested in each trial which has only been practical in captive studies of marine mammals. If dose is gradually increased, behavioural observations should be collected at a temporal scale that allows for assigning the correct dose to a response. The stimulus received by the subject should be measured (or

modelled) at the location and time of each response observation (Pater et al., 2009). It is currently not clear which property of the stimulus truly triggers a behaviour response in most situations (Ellison et al., 2012; Miller et al., 2014; Southall et al., 2007), so the dose should be quantified in different exposure metrics (or, "dose metrics") such as sound pressure level (SPL), sensation level, cumulative sound exposure level (SEL_{cum}), signal-to-noise ratio (S/N) and source-whale distance.

Dose-response studies with marine mammals have applied both quantitative and qualitative analytical methods to identify an animal's response, i.e. changes in behaviour that fall outside of baseline variation (DeRuiter et al., 2013; Miller et al., 2012). Such methods have been applied to univariate or multivariate behavioural data (e.g. measurements of vocal, movement or social parameters), resulting in dose-response functions that either describe one type of response (Antunes et al., 2014; Miller et al., 2014) or a combination of response types (Harris et al., 2015b; Sivle et al., 2015b; Williams et al., 2014). Changes in behaviour might occur in a fixed temporal order; thus, it may be possible to determine a different threshold for each response type (e.g. Blackwell et al., 2015). In general, dose-response functions predict only the onset of response and do not account for its duration or magnitude; however, severity-specific dose-response functions for different response types judged to be of similar biological importance (Harris et al., 2015b) and dose-severity functions (Houser et al., 2013a; Houser et al., 2013b) have also been reported. Harris and colleagues (2015) provided more details on the methodological considerations behind dose-response studies.

Acoustic variables are often the variables of interest in dose-response studies; however, an animal's response can also be influenced by other factors relating to the environmental or individual context of the exposed animal (Radford et al., 2016; Richardson et al., 1995). These factors therefore may increase the levels of unexplained variation in response thresholds. Examples of such contextual variables are the motivational or behavioural state of the animal (Goldbogen et al., 2013b; New et al., 2013b), personality (Sih et al., 2004), group composition (McCauley et al., 2000), density and behaviour of prey (Nowacek et al., 2011), level of ambient noise (Dunlop et al., 2013; Kastelein et al., 2011a) and relative motion of the source (Wartzok et al.,

2003). Because free-ranging animals are constantly exposed to sounds in their natural environment, the novelty of the stimulus and the animal's previous experience with that stimulus should also be important (Reichmuth, 2007). Contextual variables can be included in dose-response analyses to generate context-specific dose-response functions (Harris et al., 2015b); however, information about these covariates that may cause gradations in responsiveness is often missing, in particular for studies of free-ranging animals.

In this chapter, I used data from experiments in which tagged humpback whales were exposed to 1.3-2 kHz upsweep signals (Kvadsheim et al., 2015; Sivle et al., 2015b). These experiments were designed with two overall goals in mind: to test the effectiveness of ramp-up for naval sonar (which is addressed in Chapter 5), and to investigate all observed behavioural responses (i.e. vocal, movement, social) elicited by the sonar exposure. During the experiments, the received sonar dose was gradually escalated by having the source vessel move directly towards the whale and transmitting a ramp-up during the initial stage of the approach. This design was aimed at identifying the onset of response in the tagged animals. (Maximum levels of ~175 dB re 1 μPa for SPL and 180 dB re 1 μPa^2 s for SEL_{cum} in case of no response were aimed for). The methodology was similar to that used by Miller et al. (2014) and Antunes et al. (2014), with the primary difference being that the duration of the exposure session was shortened to 10 minutes to be able to assess sonar ramp-up effectively (Chapter 5; von Benda-Beckmann et al., 2014). The rapid dose escalation in relatively few individuals is analogous to accelerated titration methodologies developed to study the benefits and risk of medical treatments during phase-I clinical trials (Simon et al., 1997).

The two primary objectives of the current study were to 1) obtain dose-response functions for two of the response indicators investigated by Sivle et al., (2015) and 2) investigate the potential effects of the animal's behavioural state prior to response and its short-term exposure history on the response probability. Onset of horizontal avoidance and cessation of feeding were selected because they were two of the most common response types observed (Sivle et al., 2015) and because these types of behavioural change have clear potential to translate into significant effects on vital rates

if the exposure duration would be longer and the response remains constant. The analysis consisted of three parts: 1) identifying onset of avoidance and cessation of feeding, 2) measuring the response thresholds using four dose metrics (SPL, SEL_{cum}, S/N and source-whale distance), and 3) parameterising the multilevel Bayesian model described by Miller et al. (2014) using these thresholds to construct independent dose-response functions for each response type. The fine-scale movement tracks constructed in Chapter 2 were used during the identification of the avoidance responses. Additionally, results were compared to other dose-response functions published to date and differences with the recently proposed method of Harris et al. (2015) were investigated. The results of this study could have direct management implications for navies and other noise producers, but should also be relevant to the wider field of risk management of anthropogenic stressors in the marine environment.

METHODS

CEE procedure

Solitary whales or associated pairs of whales were tagged in waters north of Norway, either around Bear Island (74°N, 19°E) or West of Spitsbergen (78°N, 11°E). Whales were tagged with Fastloc-GPS loggers (F2G 134A, Sirtrack, New Zealand) that were mounted on top of multi-sensor DTAGs (version 2; Johnson and Tyack, 2003). The DTAGs recorded sound at a sample rate of 96 kHz with 16 bits resolution. The DTAGs also recorded pressure, three-dimensional (3D) acceleration and 3D magnetic field strength at a sample rate of 50 Hz. After the tags were attached, a focal follow of the tagged animal(s) was performed by visual observers from an 8-m motorised vessel. Detailed information about the data loggers and focal follow protocols can be found in Chapter 2.

Three types of exposure session were conducted as part of the experimental protocol: *No-Sonar* control sessions (n=11), *RampUp* sessions (n=18), and *No-RampUp* sessions (n=2) (Table 3.1). During *RampUp*, the source vessel approached at a speed of 4 m/s on a predetermined straight course directly towards an intercept point with the focal whale. The exposure session started with a 5-minute ramp-up, which was initiated at a distance of 1.3 km from the whale. Full-power transmissions were then continued for another 5

minutes. Each exposure session was conducted using two intercept calculators to advise the experimental coordinator because the start time of the session and line-of-approach depended upon a prediction of the future track of the whale (Kvadsheim et al., 2011).

The source vessel was the 55-m R/V HU Sverdrup II, which towed a prototype naval sonar source (Socrates II, TNO, The Netherlands; Kvadsheim et al., 2011) using 250 or 300 m of tow cable. The sonar source was positioned at a depth of about 50 m and transmitted one 1.3-2 kHz hyperbolic frequency-modulated upsweep every 20 s. During the 5-minute ramp-up period, the single-pulse source level (SL) increased according to the following sequence: 152, 168, 180, 187, 192, 196, 199, 202, 204, 206, 208, 210, 211, 213 and 214 dB re 1 μPa m. This ramp-up duration and transmission scheme was selected as the optimal ramp-up based upon a quantitative assessment of hearing injury risk (Kvadsheim et al., 2011; von Benda-Beckmann et al., 2014; see discussion in Chapter 5 for how this specific ramp-up scheme was selected). The SL of all transmissions during the 5-minute full-power period was 214 dB re 1μPa m. Sonar pulses were shorter in duration during ramp-up (0.5 s) than during full-power periods (1 s) to minimise SEL_{cum} while at the same time maintaining the expected perceived loudness of a 1-s duration pulse (Johnson, 1968; Kastelein et al., 2010).

Other session types followed the same navigational protocol as *RampUp*, but had no transmissions during the first 5 minutes. *No-RampUp* sessions had only full-power pulses during the second 5 minutes while *No-Sonar* sessions had no transmissions during the entire 10-minute period. *No-Sonar* control sessions were conducted to control for the potential effects of the vessel itself on the animals' behaviour. These control sessions were always conducted first, before the two sonar sessions, to avoid sensitisation to the sound stimulus from the vessel (Table 3.1). Up to three exposure sessions, each separated by 1 hour or more, were performed with the same whale or pair during an experimental cycle (Table 3.1).

Identification of behavioural responses

The UTC times that the humpback whales stopped feeding were reported by Sivle et al. (Sivle et al., 2015b). These authors identified lunge feeding events from sudden drops in

acoustic flow noise around the tag using a published detection algorithm (Simon et al., 2012). Feeding was considered to have ceased when the animal reached the depth of previous lunges or, if the animal was already ascending, when the next dive began. The depth measurements by the DTAG are very accurate (±1 m), thus these change points in dive profile are generally easy to identify by eye.

Sivle et al. (2015) also scored onset of avoidance; however, their judgements where based on relatively course tracks generated from Fastloc-GPS and visual position fixes. Measurement errors in these data made it sometimes difficult to judge whether or not an avoidance response had occurred. Therefore, for this thesis chapter, a similar analysis was conducted to identify onset of avoidance in the fine-scale horizontal tracks (Chapter 2). Attraction to the source was also scored but this information, along with all scores for *No-RampUp* sessions, was only used for the analysis into the effectiveness of ramp-up (Chapter 5).

Two expert groups of 2 persons each independently scored all the exposure sessions for behavioural changes using the fine-scale horizontal tracks and then reached consensus in a joint meeting. One group consisted of Dr. C. Curé (CEREMA, France) and the author; the other consisted of Dr. F. Visser (Leiden University, the Netherlands) and M. Roos (University of St Andrews). Each group was familiar with the experiments and with the lower resolution track data, but not with the fine-scale tracks. Scorers were allowed to use pre-exposure baseline data to make their judgement, because that allowed them to assess whether a behavioural change was an actual response to the sonar and/or the ship, and not just a normal and coincidental change in behaviour.

Both expert groups had scored the same 9 avoidance responses (100% agreement), and the independent scores for attraction agreed by 75% (3/4 responses) before consensus. Seven of these avoidance responses were also identified in the earlier study with the same whales (Sivle et al., 2015b). One avoidance response (mn12_179a, exposure session 2; Appendix II, Fig. A23) that was earlier not identified occurred entirely under water; the other (mn11_165, exposure session 3; Appendix II, Fig. A17) was a brief response that resulted in relatively minor horizontal displacement.

Response thresholds from the focal whale were mostly used because the movements of two associated animals are generally not independent. The only exception was exposure session 7-1, in which whale mn12_170b ceased feeding just prior to the start of ramp-up. To be precautionary in terms of impact to the whales, the cessation of feeding observed in non-focal whale mn12_170a was used as a response threshold for this session (Table 3.2).

Calculation of sonar dose

To obtain response thresholds for the dose-response analysis (Table 3.2), each behavioural response time was matched to the sonar dose received up to that point by the animal in the session. Sonar dose was described using three types of sound level, SPL, SEL and S/N, and one geometric quantity, source-whale distance. These four dose metrics quantify somewhat different aspects of the exposure. All four were considered to be potentially relevant as there is currently little scientific understanding about the causal links between acoustic dose and behavioural responses (Southall et al., 2007).

The acoustic dose metrics were calculated from the DTAG audio recordings. Digital units were converted to pressures using tag-dependent calibration values based on calibration measurements made 1 to 2 months before the experiments (Appendix III). A time-weighted SPL was measured over a 200-ms RMS averaging window for each recorded sonar pulse. This SPL was calculated by summation of the power in the third-octave bands between 1 and 40 kHz. Third-octave bands in which the signal level exceeded the noise level just prior to the pulse by less than 10 dB were excluded from this analysis. The received SPL_{max} was then taken as the maximum SPL over all pulses in the session or, if a response was scored, over all pulses that were transmitted before the response. This approach was taken to reduce the potential effects of body shielding (Wensveen, 2012). The SPL_{max} assigned to a behavioural response generally corresponded to the final pulse that was received by the animal before the onset of response as the experimental design included a rapid escalation of dose at the location of the whale.

Although the source produced harmonic energy outside of the frequency band of the sonar, this energy did not significantly contribute to the unweighted broadband SPL_{max} or SEL_{cum} (Wensveen, 2012). The third-octave bands above 2 kHz were only part of the analysis because the custom analysis program had a built-in option for applying frequency-dependent weightings. SEL_{cum} was calculated over all sonar pulses in the session or over all pulses before a response. The third-octave bands used for the calculation of SEL_{cum} were the same as used for SPL_{max}.

Background noise levels in the sonar band were combined with the SPL_{max} to obtain the S/N (Table 3.2). The noise separation method of von Benda-Beckmann et al. (2016) was applied to measure background noise levels in the tag recordings (for examples, see Figs. 3.1-3.3). This method calculates the coherence of the sound field measured on the closely spaced hydrophones to identify periods that are dominated by flow noise (i.e. flow creates an uncorrelated sound field). The SPL of the correlated part of the sound field was accepted as background noise level when it exceeded the SPL of the uncorrelated part of the sound field by 6 dB or more (von Benda-Beckmann et al., 2016). The analysis was conducted in 200-ms bins to match the (downsampled) time resolution of the DTAG sensor data. The noise separation method showed that in 79% of the sonar pulses analysed, the noise measurements in a 1-s window preceding the pulse were not affected by flow noise. In these cases the background noise level was calculated from that 1-s window (Fig. 3.3). When the noise level could not be estimated reliably (e.g. because of flow noise or surface noise), the window preceding a nearby pulse was used (Fig. 3.1). Echoes of the previous transmitted sonar signal sometimes dominated the noise level (Fig. 3.2), especially in case of the SPL_{max} of the session; thus, for these pulses, the reported S/N was in fact a signal-to-reverberation ratio.

Three experiments (IDs 1, 3 and 6) were conducted using a tag that had only one hydrophone, which precluded the use of the noise separation method for these experiments. Although aural and visual inspection (e.g. Fig. 3.2) suggested little influence of flow noise on the noise levels calculated using the single hydrophone, some of the S/Ns may be underestimated.

The sound source was assumed to closely follow the path of the vessel at regular tow speeds and turning angles. Therefore, the GPS positions of the source vessel and the depth of the source were used to estimate the 2D positions of the source by applying a correction for the time delay caused by the length of the deployed tow cable. The distances between the 3D positions of the source and the whale were calculated throughout the exposure session to obtain the minimum distance that had occurred in the session and at the start of the response.

Fitting the dose-response functions

The response thresholds during *RampUp* sessions were fitted to a set of multilevel Bayesian models to estimate the dose-response functions of the humpbacks. Benefits of using a Bayesian approach are that it facilitates a fuller representation of the uncertainties in the model parameters and is robust to small sample sizes, such as those typically collected in sonar dose-escalation studies. Further, it allows for the inclusion of prior information about model parameters and measurement uncertainty. The Bayesian model framework used in this chapter was developed by Miller et al. (2014) and is summarised below.

The process model consisted of a hierarchical organisation of variables and constants which occurred at the level of the population, individual or session. At the individual level, the expected response threshold μ_i for a behavioural response of whale *i* was assumed to follow a truncated normal distribution:

$$\mu_i \sim TN(\mu, \phi, L, U), \quad (1)$$

where μ is the mean threshold of the population, and ϕ is the between-whale s.d. in thresholds. The truncation allows for the implementation of biologically reasonable bounds on the dose-response functions. For the acoustic dose metrics, the lower bound *L* represented the dose at which the 1.3-2 kHz sonar signals were assumed to be barely detectable by the animal in the lowest sea state conditions (i.e. SPL_{max} = 60 dB re 1 μPa, SEL_{cum} = 60 dB re 1 μPa^2 s, and S/N = 0 dB (Clark and Ellison, 2004). Upper bound *U* represented a very high dose at which all individuals were assumed to have initiated a

response (SPL_{max} = 200 dB re 1 μPa, SEL_{cum} = 200 dB re 1 μPa^2 s, and S/N = 120 dB). For source-whale distance, *L* and *U* were set to 0 and 2.5 km, respectively.

The expected response threshold μ_{ij} for whale *i* in sonar session *j* was allowed to vary with one or two fixed effects:

$$\mu_{ij} = \mu_i + \beta_1 I(order)_{ij} + \beta_2 I(state)_{ij} \quad (2)$$

where β_1 is a parameter governing the effect of *RampUp2* relative to *RampUp1*, and β_2 is a parameter governing the effect of being in a foraging state relative to being in a non-foraging state. $I(order)_{ij}$ and $I(state)_{ij}$ are the corresponding binary indicator functions, which can take values of 0 or 1. Whales were scored to be in the foraging state if at least one lunge was present in the time interval from 10 min before the start of the session until either the onset of the avoidance or attraction response, or until the end of the session if no response was identified. Models for cessation of feeding omitted the β_2-term because this response was only possible in feeding animals. Gibbs Variable Selection (GVS; O'Hara and Sillanpää, 2009) was applied in order to assess the level of support in the data for including the β-terms in the model. The β-terms were excluded if the support was low.

Another random effect was included to control for within-animal variation in thresholds, but only if the GVS procedure had removed $\beta_1 I(order)_{ij}$. In that case the true but unobserved threshold t_{ij} for whale *i* in sonar session *j* was assumed to follow a truncated normal distribution:

$$t_{ij} \sim TN(\mu_{ij}, \sigma, L, U), \quad (3)$$

where σ is the within-animal between-session s.d. in thresholds, and *L* and *U* are the same as specified above. This approach diverged from the original model structure (Miller et al., 2014) which included both the fixed and random effect at the same time, but this was not possible here because session order was the same for each whale.

Further, each model structure also included an observation model to account for uncertainty in the measurements of dose. The observation error of measured threshold y_{ij} was assumed to be normally-distributed around the true threshold; $y_{ij} \sim N(\mu_{ij}, 2.5)$ for the acoustic metrics, expressed in dB, and $y_{ij} \sim N(\mu_{ij}, 25)$ for distance, expressed in metres. The observation error for acoustic dose was based upon the largest variation in DTAG sensitivity across the three acoustic calibrations (Appendix III). The observation error for distance was based upon the average positional uncertainly in the fine-scale tracks of the whales (20 m; Chapter 2; Appendix I) with some added uncertainty for the source position. If parameter t_{ij} was part of the model structure, it replaced μ_{ij} in the observation model.

Exposure sessions without responses were accounted for by including their unobserved response thresholds as censored data points (Plein and Moeschberger, 2003). For these sessions, the models for the acoustic metrics assumed a uniform prior probability of response that fell between the maximum observed dose and the upper truncation bound *U* (Table 3.3). The models for source-whale distance assumed a uniform prior probability of response that fell between 0 m and the minimum observed distance. Thus, sessions without observed responses were included either as "right-censored" or "left-censored" data depending on the dose metric used.

Model fitting was performed using Markov chain Monte Carlo (MCMC) algorithms in JAGS (v4.0, Plummer, 2003) through an interface with R (v3.2.2, R Core Team, 2013). Wide prior distributions were placed on all parameters (Table 3.3). A uniform prior was assigned to the average population threshold μ that covered the entire range over which all individual response thresholds were expected to fall; [*L*, *U*]. Parameters φ and σ had uniform priors that ranged from 0 to 0.25·(*U*–*L*), effectively allowing for random variation across this entire range of expected thresholds. Each covariate β was assigned a normal prior with mean 0 and a standard deviation of 0.25·(*U*–*L*). Posterior samples from two MCMC chains of 100,000 iterations were used for statistical inference after a burn-in period of 10,000 iterations was completed. Chains were thinned by a factor of

10 to reduce memory load. Mixing and convergence of the chains was assessed by visual inspection of trace plots and using the Brooks-Gelman-Rubin statistic $\hat{R}$ (Gelman et al., 1998; acceptance criterion: 1.05). Prior dose-response functions were calculated from the reduced model priors to visualise how the prior knowledge translated into dose-response function (Fig. 3.4).

Barnard's unconditional test (Barnard, 1945) was used on the response/no response contingency table using a null hypothesis on no effect of *RampUp* relative to *No-Sonar* control, for onset of avoidance and cessation of feeding individually, using the R package Barnard (v1.6; Erguler, 2015).

Comparisons with other dose-response functions

The functions generated for onset of avoidance and cessation of feeding in humpback whales were compared to 1) a dose-response function constructed using an alternate method that was applied to behavioural threshold data from the same experiments, and 2) dose-response functions for behavioural responses to sonar by other marine mammal species. The aims of these comparisons were to assess the effects of methodological assumptions behind the methods and to evaluate the potential reasons for differences between dose-response functions.

1) The alternate method was a type of a recurrent event survival analysis in which marginal stratified Cox proportional hazards models were fitted to the data (see Harris et al., 2015, for full details of the methodology). Here, this method was applied to the expert-scoring data set of Sivle et al. (2015) which included additional response types such as changes in diving and changes in group distribution. The data set was checked for errors and updated with the avoidance onset thresholds based on the fine-scale tracks (Table 3.2) that were identified in this chapter. Models were fitted only to SEL_{cum} thresholds for *RampUp* sessions. The analysis predicted the probability of response within three separate response severity strata, but, because the avoidance and change-in-feeding thresholds used in the Bayesian models all fell within the second stratum (i.e. response severities 4 to 6; Southall et al., 2007), only these specific functions were plotted. Both fixed effects in Eqn 3.2 were included initially; however, foraging state

did not meet the proportional hazards assumption (Kleinbaum and Klein, 2005) and session order was excluded during backwards model selection (Harris et al., 2015), so only the dose-response functions for models without β-terms were compared. Model fitting was performed in R using the survival package (v2.38-3; Therneau, 2015).

2) Bayesian models were also fitted to the combined avoidance threshold data for humpback whales, killer whales (Miller et al., 2014) and long-finned pilot whales (Antunes et al., 2014). Random between-species variation was not accounted for, so these models assumed that all whales came from the same population. Dose-response functions were generated from two subsets of data, one that included and one that excluded responses of killer and pilot whales during 6-7 kHz sonar sessions. These models did not control for session order or foraging state (Eqn 2) but included two fixed effects for species; killer whale relative to humpback whale (β_3) and long-finned pilot whale relative to humpback whale (β_4). The priors on the covariates were given the same values as for β_1 and β_2 (Table 3.3). The resulting species-specific dose-response functions were compared visually to published dose-response functions for the onset of behavioural responses in other marine mammal species (Houser et al., 2013a; Houser et al., 2013b; Kastelein et al., 2012; Moretti et al., 2014).

RESULTS

Eleven independent groups of 1 or 2 tagged humpback whales were subjected to 11 *No-Sonar* controls and 18 *RampUp* sessions (i.e. 10 *RampUp1* and 8 *RampUp2*; Table 3.1). The sonar dose received by the whales during *RampUp* sessions ranged between 86 and 182 dB re 1 μPa for SPL_{max}, 81 and 184 dB re 1 μPa^2 s for SEL_{cum}, –3 and 88 dB for S/N, and 0.062 and 2.2 km for source-whale distance (Fig. 3.5 and 3.6). The variability in observed responses thresholds was large; thresholds ranged (mean ± s.d.) between 102 and 179 dB re 1 μPa (146 ± 21) for SPL_{max}, 97 and 179 dB re 1 μPa^2 s (144 ± 22) for SEL_{cum}, 17 and 83 dB (51 ± 15) for S/N, and 0.21 and 1.7 km (0.98 ± 0.47) for source-whale distance (Fig. 3.5).

The observed response thresholds revealed correlations between the four dose metrics at the time of response. The dose-escalation design simultaneously increased SPL_{max} and decreased source-whale distance (Fig. 3.5A; $r = -0.74$) during the approach period (all but one response occurred during this period). A very strong positive correlation ($r = +0.98$) between SPL_{max} and SEL_{cum} indicated that SEL_{cum} was dominated by the SEL of the most recent pulse as the increase in the single-pulse SEL (and SPL) from one pulse to the next was relatively high (Fig. 3.5B). More unusual was the increase in SPL_{max} with the noise level measured just before the pulse (Fig. 3.5D; $r = +0.85$), which suppressed the variation in S/N compared to SPL_{max} (Figs. 3.5A,C). This positive correlation likely resulted from sonar reverberation levels still exceeding the ambient noise level (for pulses with high source levels) and/or from noise radiating from the source vessel (for pulses earlier in the session).

Nine avoidance responses during all 18 *RampUp*s (50%) and 5 cessations of feeding during 8 *RampUp*s in which animals were initially feeding (63%) were observed (Table 3.2; Appendix II). Whales neither avoided nor stopped feeding in 6 out of 18 *RampUp*s (33%), and both response types occurred only twice within the same session (Table 3.2). Avoidance was never observed during the 11 control sessions (Appendix II), strongly suggesting that the avoidance responses during *RampUp* were caused by the sonar stimuli and not the source vessel (Barnard's test: Wald statistic = 4.24, nuisance = 0.34, $p = 0.001$). In contrast, 2 cessations of feeding were scored for 7 *No-Sonar* control sessions in which animals were initially in a feeding state, so statistical evidence for this response type was weaker (Barnard's test: Wald statistic = 1.40, nuisance = 0.24, $p = 0.13$). However, one of these two responses was scored with low confidence by Sivle et al. (2015) because only one animal of the pair had stopped feeding.

Thresholds were fitted to multilevel Bayesian models to generate the dose-response functions. The Brooks-Gelman-Rubin statistic $\hat{R}$ and visual inspection of model diagnostic plots showed that all parameters quickly converged to stationary distributions with good mixing of the chains, with $\hat{R}$ values always falling well below the pre-set criterion of 1.05. The MCMC posterior chains for the SPL_{max} model for onset of avoidance illustrated in Figs. 3.7 and 3.8 were representative for the other models.

The posterior probabilities of the parameters in the full models that included the factors session order and foraging state are summarised in Table 3.4. Context-specific dose-response functions were generated from the full SPL_{max} model for onset of avoidance to investigate how the response probabilities varied across the four subsets in these whales (Fig. 3.9). On average, non-feeding whales during *RampUp1* had the lowest avoidance thresholds (p_{50}: 128 dB re 1 μPa), followed by non-feeding whales during *RampUp2* (p_{50}: 158 dB re 1 μPa), feeding whales during *RampUp1* (p_{50}: 177 dB re 1 μPa), and feeding whales during *RampUp2* (p_{50}: 186 dB re 1 μPa). The relatively high and precise posterior probabilities of the β parameters suggested similar differences between foraging/non-foraging states and between *RampUp1* and *RampUp2* (Fig. 3.7). These context-specific functions should be interpreted with caution, as small numbers of observed thresholds within some of the subsets and whales not responding at the maximum sonar dose resulted in very low support for inclusion of the β-terms in the final model (GVS *p*-values between 0.02 and 0.09; Table 3.4). Models for cessation of feeding were not interpreted because the subset for *RampUp2* consisted of only one observed response threshold (and 2 non-responses). There was a high level of unexplained between-whale variation in thresholds (φ) in these full models; the posterior medians for this parameter were estimated to be close to the upper bounds of their priors (Table 3.4).

Between-whale variation in the reduced models without β-terms was about half (on the scale of the s.d.) of the between-whale variation in the full models; this part of the variation was absorbed by the newly-added random effect for within-whale variation (parameter σ; Tables 3.4 and 3.5). For onset of avoidance, the sonar doses at which the response was predicted to occur in 50% of the animals (p_{50}) were 166 dB re 1 μPa (95% CI: 151-176) for SPL_{max}, 168 dB re 1 μPa^2 s (95% CI: 148-179) for SEL_{cum}, 71 dB (95% CI: 56-86) for S/N, and 578 m (95% CI: 427-843) for source-whale distance (Fig. 3.10; Table 3.5). Dose-response functions for cessation of feeding were almost identical to those for onset of avoidance, but had slightly wider 95% CIs due to greater within-animal variation.

Behavioural responses of all severities were used in the recurrent event survival analysis (Harris et al., 2015), but only the dose-response function for response severities 4-6 was plotted because all cessations of feeding and avoidance responses fell within this middle stratum. This dose-response function was similar in shape and location to the dose-response functions generated from the reduced Bayesian models; the difference in estimated expected (mean or median) response probability for a given dose never exceeded 0.2 (Fig. 3.11). Neither the Cox proportional hazards dose-response function nor the Bayesian dose-response functions ever fell outside the others' 95% confidence bounds.

Models were also fitted to a combined data set that included avoidance responses by humpback whales, killer whales and long-finned pilot whales (Miller e al., 2014; Antunes et al., 2014). The predicted differences in average response threshold between the species was in the order of 5 to 25 dB (Figs. 12A,B; see Table 3.6, for posterior summary statistics of all parameters). In particular, most of the median curve of killer whales for the combination of 1-2 kHz and 6-7 kHz thresholds fell below the 95% CIs of the other two species. This difference was reduced when the model was re-fitted to include only thresholds measured during 1-2 kHz sessions, as the thresholds of killer whales were generally lower during 6-7 kHz sessions (Miller et al., 2014). Humpback whales were slightly more responsive during sonar exposure than long-finned pilot whales (Figs 12A,B), although the differences between the curves were small. The GVS procedure indicated little statistical support in favour of the β-terms for species (GVS p-values: 0.31-0.40), so latent contextual variables or the small sample sizes may have caused any of these differences.

DISCUSSION

Sonar dose-escalation experiments and a Bayesian statistical procedure were used here to evaluate behavioural response onset in relation to three acoustic exposure metrics, distance to the source and two contextual variables. The analysis dealt appropriately with the small sample size, sessions without observed responses and prior assumptions. Similar methodology has been used by Miller et al. (2014) and Antunes et al. (2014) to describe the probability of avoidance onset in killer whales and long-finned pilot

whales, respectively. An important caveat that applies to all these studies is that, akin to in phase-I clinical trials, relatively few individuals were tested, so that results should be considered to provide only the first indication of the variation in response thresholds for free-ranging cetaceans. Ideally, the predictions made by these experiments are compared in future observational research involving more individuals exposed to operational sonar sources, e.g. by using PAM, visual surveys and satellite tag deployments closely in time and space with actual naval exercises.

Methodological considerations

The multi-variate behavioural data sets collected during the humpback whale experiments were described in detail by Kvadsheim et al. (2015). From these data sets, Sivle et al. (2015) qualitatively identified movement, social and vocal responses during control and sonar sessions and scored them to a modified version of the Southall et al. (2007) severity scale. Scorers were allowed to inspect baseline behaviour patterns to distinguish responses from normal changes in behaviour. Here, I used the cessations of feeding identified in that study, which were based upon high-resolution DTAG dive profiles and upon feeding lunges that were detected from rapid decreases in speed (Simon et al., 2012). Sivle et al. (2015) also scored onset of avoidance based on surface positions from GPS and visual sightings; however, in this chapter, the high-resolution movement tracks constructed in Chapter 2 were used because more subtle responses may have been missed in that first assessment. Indeed, all seven avoidance responses scored by Sivle et al. (2015) were confirmed here and two additional avoidance responses were scored which were not identifiable from the noisier surface positions (Table 2; see Appendix II for all whale tracks). This illustrates how behavioural changes may be missed if they occur entirely underwater (mn12_179, *RampUp1*; Fig A23) or at scales below the temporal or spatial resolution of the observations (mn11_165, *RampUp2*; Fig. A17).

Only the response thresholds that were measured during dose-escalation experiments (i.e. *RampUp* sessions; Table 3.1) were used in the Bayesian models. Comparison to the responses during *No-Sonar* sessions showed that onset of avoidance was very likely to be caused by the sonar stimulus and not by the approaching source vessel, thus

providing strong support that its dose-response function was representative for the effects of sonar. Compared to onset of avoidance, cessation of feeding was more likely to be elicited by the sonar stimulus or the source vessel. The timing of feeding cessation was also more difficult to determine precisely because humpback whales re-orient themselves and need time to process food after a lunge. Both of these effects may have added to the uncertainty in the dose-response functions for cessation of feeding.

Dose-response functions were created for four dose metrics. The exposure metric that is most often reported in marine mammal disturbance studies is the received SPL (Southall et al., 2007), but more recent studies also often report SEL (e.g. Blackwell et al., 2004). Animals might not only respond to absolute levels such as SPL and SEL but also to levels relative to the hearing threshold (Ellison et al., 2012); frequency weighting can provide a solution to this issue (Chapter 4). However, this method was not used here as the sonar band (1.3-2 kHz) fell within the most sensitive hearing range of humpback whales according to relative audiograms derived from anatomical modelling (Houser et al., 2001; Ketten and Mountain, 2014). High-frequency harmonics or sideband energy are likely part of the signals from many operational naval sonar systems (e.g. Melcón et al., 2012) and these spectral components should be considered in the assessment of sonar effects on marine mammals (Kastelein et al., 2012; Miller et al., 2012). In this study, however, the received SPLs outside of the sonar band were too low to be expected to significantly influence a weighted or unweighted broadband level metric relevant for humpback whales.

Animals might also respond to levels relative to the competing background noise (e.g. S/N and signal excess; Ellison et al., 2012). Indeed, during playbacks of ~2-kHz signals to migrating humpback whales, the received S/N at the start of the playback was a better predictor of changes in dive behaviour than the received SPL or source-whale distance (Dunlop et al., 2013). Dunlop and colleagues' (2013) used observation methods that excluded the precise identification of response onsets, although the signal level was expected to be close to the background noise level during most observed responses. In humans, acoustic annoyance of intruding sounds correlated better with absolute levels than with S/N, except if the noise level approaches to near the signal level (Richardson

et al., 1995). This is consistent with the findings of Kastelein et al. (2011), who found that a harbour porpoise reduced the intensity of behavioural responses to 6-7 kHz sweeps when signal levels were 20 dB or less above the level of a broadband masker (simulating sea states 4 or less); however, there was no difference in responsiveness at S/N = 26 and 41 dB. It is conceivable that the influence of background noise on the response thresholds in the current study also diminished with increasing SPL. Therefore, background noise may have elevated particularly the lowest absolute threshold (i.e. SPL_{max} = 102 dB re 1 μPa; S/N = 17 dB; Table 3.2).

The novel method of von Benda-Beckmann et al. (2016) was used here to assess if noise levels were affected by flow noise and it proved to be a very useful tool. Whilst flow noise often dominated the noise levels at frequencies below 1 kHz, levels at higher frequencies were generally not affected unless the speed through the water of the tagged animal was relatively high (>2 m/s). Sonar echoes and noise from the source vessel affected the underwater soundscape during the exposure sessions. As a result, the noise levels varied with depth, source-whale distance and time since previous pulse. Therefore, I chose to estimate the background noise level in a short (1 s) analysis window recorded just before the pulse.

By design, the vessel approach escalated not only the received sound levels at the focal whale but also the proximity of the source to the whale. As this parameter was controlled by the experimenter, source-whale distance was analysed as a dose metric. Effects of source-whale distance on behavioural responses have been shown in a number of studies including several studies on baleen whales (Dunlop et al., 2013; Frankel and Clark, 1998; Frankel and Clark, 2000b; Gailey et al., 2007; Maybaum, 1989). Disentangling the potential effects of range and received level is often difficult (but see Frankel and Clark, 2000) because these variables are often correlated (as in Fig. 3.5). Recent observations of the absence of responses in tagged cetaceans during incidental sonar exposures highlighted the importance of the interaction between range and received level (Cuvier's beaked whales: DeRuiter et al., 2013; sperm whales: Isojunno et al., 2016). These observations suggested that, for comparable received levels, transmissions from a distant sonar source at high source level may be less

aversive than transmissions from a nearby source at lower source level. The incidental sonar exposures were received in the beginning of the tag records, however, so the tagged whales may have habituated to the sounds before the start of the experiments.

Whales were exposed to sonar doses that covered most of the range over which responses were expected to be theoretically possible (Fig. 3.6). It is worthwhile to consider the influence of the priors (Table 3.3) and other model assumptions on the dose-response functions, in particular at levels outside the realised exposure range. For example, the model assumed that all whales initiated a response at source-whale distance = 0 m, SPL_{max} = 200 dB re 1 μPa, SEL_{cum} = 200 dB re 1 μPa^2 s and S/N = 120 dB, although these doses were never reached. There is little scientific evidence on behavioural effects from such high sound levels or close proximities. The value of this upper boundary and the way that censored data were included had some influence on the dose-response functions because of the relatively high number of sessions without observed response. The censoring method was unbiased in the sense that censored response thresholds were assumed to be distributed uniformly between the realised maximum dose and the assumed upper boundary.

Only one response occurred very early in the session and none directly after the first sonar pulse (Fig. 3.6), which suggested that the low end of the dose-response functions were not severely affected by the lack of low-level exposures (there might have been some effect of noise level, however, as mentioned earlier). For the acoustic metrics, the minimum value that a threshold could take (i.e. parameter *L*; Eqn 3.1) represented the dose at which a 1.3-2 kHz sonar signal was assumed to be barely detectable to the animal in very low ambient noise conditions (Clark and Ellison, 2004). Eqn 4.5 describes how such theoretical detection thresholds can be calculated from average wind-driven noise levels given sufficient available information (see Fig. 4.9 for an example using the harbour porpoise). Because essentially nothing is known about the masked hearing abilities of baleen whales, these minimum values are only approximations for humpback whales. Regardless of the derivation of the dose-response functions, the application of these functions to predict responses in other situations

should always be done carefully and with consideration of the audibility of signals in quiet and noisy conditions.

Good correspondence was generally found between the dose-response functions generated by the Bayesian models and Cox proportional hazards model for severities 4-6 (Fig. 3.11). The method proposed by Harris et al (2015) and applied by Sivle et al. (2015) to humpback whale data allows for the inclusion of different response types (e.g. avoidance, vocal responses) and does not require response types to be selected *a priori*, as such is the case in the Bayesian model (although this judgement is made implicitly via the Southall et al. (2007) severity scale). In its current form, the method acknowledges that responses may not be equally severe; that is, some response types and responses of longer duration are assumed to be more biologically-significant than others. The Cox proportional hazards dose-response function predicted a slightly higher probability of response at SEL_{cum} below ~150 dB re 1 μPa^2 s (Fig. 3.11) compared to the Bayesian functions. This difference was likely the result of other response types occurring before onset of avoidance and cessation of feeding as only the first occurrence of each severity level is part of the input data.

While the Cox proportional hazards dose-response function approximated the raw threshold data more closely, the Bayesian models had stronger distributional assumptions and therefore produced smoother curves (Fig. 3.11). These assumptions also allowed for prediction outside of the range of observed response thresholds and for the incorporation of measurement uncertainty. The use of a single response type made the output of the Bayesian function easier to interpret; however, in theory, a different function for each response type could also be generated using the method of Harris et al. (2015) by stratifying the onset thresholds by response type instead of response severity. Although both methods allowed for the incorporation of covariates, the model selection via GVS appeared to be perhaps less powerful than model selection based on Akaike information criterion that was employed by the Cox proportional hazards-based approach. Thus, while both approaches produced comparable results in this case, some differences in their outputs were present here and should be expected. Which is the most

appropriate method will depend upon the goal of the analysis and the quality of the input data.

Factors affecting behavioural responses

This study uncovered dose-response functions for two functionally distinct behavioural response types, i.e. cessation of feeding and onset of avoidance, exhibited by humpback whales exposed to naval sonar. Functions were very similar between cessation of feeding and onset of avoidance despite the differences in sample size. Both response types occurred only twice within the same session (Table 3.2); that is, in 25% of sessions in which the whale was initially in feeding state. This suggested that horizontal displacement and reductions in feeding occur at similar disturbance levels in humpback whales, even if they were generally not the components of one overall response. This prediction is consistent with the severity scale of Southall et al. (2007), which assigned the same level of importance to these response types.

Onset of avoidance and cessation of feeding were modelled here due to their potential to translate into biologically significant effects. However, observations of no response should not be considered as evidence of no impact as whales may only respond physiologically and not behaviourally or by changing behaviour that is not observed (Tougaard et al., 2015b). For example, at received SPLs of 175 dB re 1 μPa or higher, the probability of long-finned pilot whales using a synchronous surfacing strategy to reduce received levels exceeded the probability of this species avoiding the sonar source horizontally (Antunes et al., 2014; Wensveen et al., 2015a).

A comparison of dose-response functions ranked the humpback whale as being less responsive to sonar signals on average than most other species tested (Figs 3.12C; Houser et al., 2013a; Houser et al., 2013b; Kastelein et al., 2012; Moretti et al., 2014). Large differences in response thresholds were apparent across the studies, especially at the lower probabilities of response. These differences were likely the result of a complex mix of factors, e.g. the experimental setting (wild vs. laboratory), type of response, hearing sensitivity and temperament of the species, acoustic and non-acoustic properties of the source, group composition, age and sex distribution of the animals,

prior experience with the stimulus. Below, I discuss the behavioural response thresholds and dose-response functions derived in this chapter in the broader context of what is known about marine mammal behavioural responses to disturbance.

Acoustic variables

The response threshold SPLs observed here (Table 3.2) fell within or just above the range of values reported in other studies with humpback whales. The dose-response functions were relatively shallow (Figs. 3.9 and 3.10), which reflected the large variation in thresholds that is typical for studies with free-ranging marine mammals (Fig 3.12). Southall et al. (2007) summarised the responses of large baleen whales to nonpulsed sounds as no (or very limited) responses between 90 and 120 dB re 1 μPa and an increasing probability of avoidance or other behavioural effects in the 120 to 160 dB re 1 μPa range. This global conclusion was based on responses to sonar signals but also on responses to very different stimuli such as broadband vessel noise (e.g. Baker et al., 1983).

Responses of humpbacks to tonal signals are particularly worth mentioning. In a study that aimed at behaviourally estimating auditory detection thresholds, Frankel et al. (1995) found rapid approach responses to playbacks of continuous, frequency-modulated (FM) synthetic sound (0.01-1.4 kHz) in 4% of humpback whale pods at a median received SPL of 113 dB re 1 μPa (lowest threshold: 106 dB re 1 μPa). At similarly low received SPLs (88-110 dB re 1 μPa), reductions in humpback whale song were observed in response to FM upsweeps (0.4-1.0 kHz band) produced by a high-power fisheries sonar (Risch et al., 2012). Maybaum (1989, 1993) reported that the linearity of humpback whale tracks increased with estimated received SPLs over a range of ~110-140 dB re 1 μPa during controlled exposures to FM sweeps (3.1-3.6 kHz) transmitted by a Massa M-1002 sonar or a transducer with a lower source level. Humpback whales were also reported to increased their song lengths at maximum received SPLs of ~110-150 dB re 1 μPa in response to exposure to low-frequency active sonar signals (0.1-0.5 kHz band) at a reduced source level (Biassoni et al., 2000; Miller et al., 2000). Observations of no response were common in most of these above studies and whales were not exposed to SPLs above 160 dB re 1 μPa. Therefore, there is

reasonable correspondence between the SPLs reported elsewhere and the threshold SPLs of the current study, although most of the levels reported elsewhere fall at the low end of the posterior dose-response functions (Fig. 3.10; p_{50} value = 166 dB re 1 μPa).

Some of these differences in threshold SPLs across studies were likely caused by the spectral or temporal characteristics of the exposure signals. The potential importance of the frequency spectrum is exemplified in Kastelein et al. (2012), who derived three dose-response functions for brief changes in direction and speed for a captive harbour porpoise (Fig 3.12C). The functions exhibited clear differences in received SPLs between 1-2 kHz upsweeps without harmonics, 6-7 kHz upsweeps without harmonics and 1-2 kHz upsweeps with strong harmonics (a signal with maximum sensation level at ~12 kHz). The frequency spectra of sonar signals may also have affected the behavioural responses of free-ranging sperm whales during controlled sonar exposures (Isojunno et al. 2016). Consistent changes in foraging behaviour were observed in response to 1-2 kHz sweeps, but not in response to 6-7 kHz sweeps (both signals had weak harmonics). However, the 15 dB difference in source level between the two signal types may also have affected the responses of the sperm whales.

If information about loudness perception or absolute hearing sensitivity is available for a species, effects of the spectral characteristics of the signal can be accounted for by using a frequency-weighted broadband SPL as the dose metric (Chapter 4). Such an approach was taken by Miller et al. (2014) for avoidance responses of free-ranging killer whales to naval sonar signals; however, the evidence for a potential effect of signal frequency in threshold was not conclusive. Here, the dose-response functions of these killer whales were recreated by fitting Bayesian models to a combined data set with long-finned pilot whales (Antunes et al., 2014) and humpback whales (this study; Fig. 3.12A,B). Refitting the model without the 6-7 kHz data caused an upward shift in the dose-response function for killer whales but not for long-finned pilot whales, confirming some effect of sonar frequency in the first species but not the latter.

Contextual variables

High levels of between-animal and/or within-animal variability in response thresholds were estimated here (Tables 3.4 and 3.5), indicating that thresholds depended upon unidentified contextual variables (e.g. psychological, physiological, social or environmental factors). The small sample size and large variability in thresholds resulted in very little statistical support for the two contextual variables formally investigated, i.e. session order and feeding state. Therefore, firm conclusions about these effects cannot be drawn; however, the estimated effects were large and there were some other interesting trends present in the data set, which I will describe below.

Avoidance thresholds were generally lower during the second sonar session (*RampUp2)* one hour after the first sonar session (*RampUp1*). This difference was most pronounced for animals in a non-feeding state (Fig. 3.9). These response thresholds may thus have varied as a function of the novelty of the stimulus or short-term habituation. Separate analyses in Sivle et al. (2015) and in Chapter 5 supported the possibility of less severe and fewer responses by humpbacks during the second session. Habituation can occur only in a specific exposure SPL range; for example, the dose-response functions for various behavioural responses of captive bottlenose dolphins indicated an effect of trial number for SPLs below 175 dB re 1 μPa, but not at higher levels (Fig. 3.12C; Houser et al., 2013a). A decrease in behavioural responsiveness was not observed for California sea lions that were tested using the same experimental design (Houser et al., 2013b). While animals in human care should not be considered as directly representative of wild animals, these studies indicated that the potential for habituation (and the time scale over which it occurs) depends upon received level and species' temperament (Lowry et al., 2013).

Houser et al. (2013a, 2013b) also found that responses depended upon the age distribution of the animals. This effect was particularly clear for the sea lions, where animals of 1 or 2 years old showed stronger behavioural responses than older animals, which disproportionately affected the shape of the dose-response function (Fig. 3.12C; Houser et al., 2013b). Other aspects of the composition of the social group (e.g. group size, calf presence) can also affect the responses of whales to stressors (Cantor et al.,

2010; Dunlop et al., 2013; Tyack, 1983). The dataset used here included three potential mother-calf pairs, defined as duos that were composed of an adult and a smaller-sized individual which remained closely associated with each other throughout the tracking record (Curé et al., 2015). One of these calves was substantially smaller than the other two (Fig. 5.3). The mother of this particular calf (mn12_180ab; Table 3.1; the calf was not tagged) was more responsive; it made an unusually deep dive during each sonar session and it was the only whale in the data set that showed an avoidance response during both sonar sessions (Appendix II). In addition, its avoidance threshold for *RampUp2* was 38 dB lower than its avoidance threshold for *RampUp1* (Table 3.2), suggesting sensitisation of this whale and/or its untagged calf.

There was also evidence that the behavioural state of the animals (feeding/non-feeding) influenced the avoidance thresholds (Fig. 3.9). In both sonar sessions, whales in a non-feeding state responded earlier and with lower thresholds than whales in a feeding state. Very similar effects of behavioural state on responses to naval sonar have been described in blue whales (Goldbogen et al., 2013b). Blue whale responsiveness to sounds was reduced when whales were shallow-feeding (max dive depth <50 m) compared to non-feeding and deep-feeding whales. There was no evidence of a useable depth criterion here but most feeding dives were shallower than 50 m (Appendix II). Behaviour is an expression of the underlying motivational state of an animal (McFarland and Sibly, 1975; New et al., 2013b), so one might predict that hungry whales may tolerate sonar exposure for longer because of the benefit to the animal's fitness. An example of this was whale mn12_178a in experiment 9, which was also the most emaciated animal in the data set (for a photo, see Kvadsheim et al. 2015, p111). This animal fed continuously throughout the tag record and exhibited no cessation of feeding and only one avoidance response (classified as 'brief'; Sivle et al., 2015) at an average received SPL of 143 dB re 1 μPa (Table 3.2; Appendix II, Fig. A11).

Implications for management

Dose-response functions for the onset of avoidance and cessation of feeding were successfully established for humpback whales exposed to 1.3-2 kHz naval sonar. Comparing these results to response thresholds reported for humpback whales and other

marine mammal species suggested somewhat lower responsiveness in humpback whales on average (Fig. 3.12). However, the relatively shallow slopes of the functions indicated that some individuals may respond at relatively low received levels that could be equivalent to long distances from higher-power operational sonar sources. It is unlikely that the short behavioural responses observed during these experimental exposures affected the health of the individuals, but responses such as horizontal displacement and disruptions of feeding have a clear potential to translate into population-level impacts if sonar exposure is more prolonged, more intense and cumulated with other stressors.

While dose-response functions are presented here for four dose metrics that are potential predictors of responses, these functions represent responsiveness in the context of the conducted experiments. The degree to which these dose metrics can be used to extrapolate the response probability to other situations is uncertain. Theoretically, a single dose-response relationship is only truly valid if all other dose metrics are irrelevant, but there are clear discrepancies between the predictions made by the different functions. For example, if only distance would be important then essentially no responses would occur at 2 km from the source. However, this distance translates to about 150 dB re 1 μPa under spherical spreading conditions, a value at which response onset in a significant proportion of animals is predicted by the SPL_{max} dose-response function. The relative importance of these different dose metrics represents an important gap in scientific knowledge which bears relevance to extrapolating results to sonars of higher source levels and establishing noise criteria for behavioural disturbance.

REFERENCES

Ainslie, M. A. (2010). Principles of Sonar Performance Modeling. Chichester, UK: Springer-Praxis.

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

Atkinson, S., Crocker, D., Houser, D. and Mashburn, K. (2015). Stress physiology in marine mammals: how well do they fit the terrestrial model? J. Comp. Physiol. B 1–24.

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. National Marine Fisheries Service, National Marine Mammal Laboratory.

Balcomb, K. C. and Claridge, D. E. (2001). A mass stranding of cetaceans caused by naval sonar in the Bahamas. Bahamas J. Sci. 5, 2–12.

Barnard, G. A. (1945). A new test for 2x2 tables. Nature 156, 177.

Bejder, L., Samuels, A., Whitehead, H. and Gales, N. (2006). Interpreting short-term behavioural responses to disturbance within a longitudinal perspective. Anim. Behav. 72, 1149–1158.

Bejder, L., Samuels, A., Whitehead, H., Finn, H. and Allen, S. (2009). Impact assessment research: use and misuse of habituation, sensitisation and tolerance in describing wildlife responses to anthropogenic stimuli. Mar. Ecol. Prog. Ser. 395, 177–185.

Biassoni, N., Miller, P. J. O. and Tyack, P. L. (2000). Preliminary results of the effects on SURTASS-LFA sonar on singing humpback whales. Woods Hole, Massachusetts: WHOI.

Blackwell, S. B., Lawson, J. W. and Williams, M. T. (2004). Tolerance by ringed seals (*Phoca hispida*) to impact pile-driving and construction sounds at an oil production island. J. Acoust. Soc. Am. 115, 2346–2357.

Blackwell, S. B., Nations, C. S., McDonald, T. L., Thode, A. M., Mathias, D., Kim, K. H., Greene, C. R. and Macrander, A. M. (2015). Effects of airgun sounds on

bowhead whale calling rates: evidence for two behavioral thresholds. PLoS One 10, e0125720.

Cantor, M., Cachuba, T., Fernandes, L. and Engel, M. H. (2010). Behavioural reactions of wintering humpback whales (*Megaptera novaeangliae*) to biopsy sampling in the western South Atlantic. J. Mar. Biol. Assoc. UK 90, 1701–1711.

Clark, C. W. and Ellison, W. T. (2004). Potential use of low-frequency sounds by baleen whales for probing the environment: evidence from models and empirical measurements. In Echolocation in Bats and Dolphins (ed. Thomas, J. A., Moss, C. F. and Vater, M.), pp. 564–582.

Clark, C., Ellison, W. T., Southall, B. L., Hatch, L., Van Parijs, S. M., Frankel, A. and Ponirakis, D. (2009). Acoustic masking in marine ecosystems: intuitions, analysis, and implication. Mar. Ecol. Prog. Ser. 395, 201–222.

Cox, T. M., Ragen, T. J., Read, A. J., Vos, E., Baird, R. W., Balcomb, K., Barlow, J., Caldwell, J., Cranford, T., Crum, L., et al. (2006). Understanding the impacts of anthropogenic sound on beaked whales. J. Cetacean Res. Manag. 7, 177–187.

Curé, C., Doksaeter 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.

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

DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone, E. A., Friedlaender, A. S., Joseph, J. E., Moretti, D., Schorr, G. S., et al. (2013). First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. Biol. Lett. 9, 20130223.

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.

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). R Package “Barnard.”

Fernández, A., Edwards, J. F., Rodríguez, F., Espinosa de los Monteros, A., Herráez, P., Castro, P., Jaber, J. R., Martín, V. and Arbelo, M. (2005). “Gas and fat embolic syndrome” involving a mass stranding of beaked whales (Family *Ziphiidae*) exposed to anthropogenic sonar signals. Vet Pathol 42, 446–457.

Finneran, J. J. (2015). Noise-induced hearing loss in marine mammals: A review of temporary threshold shift studies from 1996 to 2015. J. Acoust. Soc. Am. 138, 1702–1726.

Frankel, A. S. and Clark, C. W. (1998). Results of low-frequency playback of M-sequence noise to humpback whales, *Megaptera novaeangliae*, in Hawaii. Can. J. Zool. 76, 521–535.

Frankel, A. S. and Clark, C. W. (2000). Behavioral responses of humpback whales (*Megaptera novaeangliae*) to full-scale ATOC signals. J. Acoust. Soc. Am. 108, 1930–1937.

Frankel, A. S., Mobley Jr, J. R. and Herman, L. M. (1995). Estimation of auditory response thresholds in humpback whales using biologically meaningful sounds. In Sensory Systems of Aquatic Mammals (ed. Kastelein, R. A., Thomas, J. A., and Nachtigall, P. E.), pp. 55–79. Woerden, The Netherlands.

Gailey, G., Wursig, B. and McDonald, T. L. (2007). Abundance, behavior, and movement patterns of western gray whales in relation to a 3-D seismic survey, Northeast Sakhalin Island, Russia. Environ. Monit. Assess. 134, 75–91.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1998). Bayesian Data Analysis. (ed. Chatfield, C. and Zidek, J. V.) London, UK: Chapman & Hall.

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.

Harris, C. M., Sadykova, D., DeRuiter, S. L., Tyack, P. L., Miller, P. J. O., Kvadsheim, P. H., Lam, F. P. A. and Thomas, L. (2015). Dose response severity functions for acoustic disturbance in cetaceans using recurrent event survival analysis. Ecosphere 6, art236.

Houser, D. S., Helweg, D. A. and Moore, P. W. B. (2001). A bandpass filter-bank model of auditory sensitivity in the humpback whale. Aq. Mam. 27, 82–91.

Houser, D. S., Martin, S. W. and Finneran, J. J. (2013a). Exposure amplitude and repetition affect bottlenose dolphin behavioral responses to simulated mid-frequency sonar signals. J. Exp. Mar. Bio. Ecol. 443, 123–133.

Houser, D. S., Martin, S. W. and Finneran, J. J. (2013b). Behavioral responses of California sea lions to mid-frequency (3250-3450 Hz) sonar signals. Mar. Environ. Res. 92, 268–278.

Isojunno, S., Curé, C., Kvadsheim, P., Lam, F., Tyack, P., Wensveen, P. and Miller, P. (2016). Sperm whales reduce foraging effort during exposure to 1-2 kHz sonar and killer whale sounds. Ecol. Appl. in press,.

Johnson, C. S. (1968). Relation between absolute thresholdand duration-of-tone pulses in the bottlenosed porpoise. J. Acoust. Soc. Am. 43, 757–763.

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.

Kastelein, R. A., Hoek, L., de Jong, C. A. and Wensveen, P. J. (2010). The effect of signal duration on the underwater detection thresholds of a harbor porpoise (*Phocoena phocoena*) for single frequency-modulated tonal signals between 0.25 and 160 kHz. J. Acoust. Soc. Am. 128, 3211–3222.

Kastelein, R. A., Steen, N., de Jong, C., Wensveen, P. J. and Verboom, W. C. (2011). Effect of broadband-noise masking on the behavioral response of a harbor porpoise (*Phocoena phocoena*) to 1-s duration 6-7 kHz sonar up-sweeps. J. Acoust. Soc. Am. 129, 2307–2315.

Kastelein, R. A., Steen, N., Gransier, R., Wensveen, P. J. and de Jong, C. A. (2012). Threshold received sound pressure levels of single 1-2 kHz and 6-7 kHz up-sweeps and down-sweeps causing startle responses in a harbor porpoise (*Phocoena phocoena*). J. Acoust. Soc. Am. 131, 2325–2333.

Ketten, D. R. and Mountain, D. C. (2014). Inner ear frequency maps: First stage audiograms of low to infrasonic hearing in mysticetes. In 5th International Meeting on the Effects of Sound in the Ocean on Marine Mammals, Amsterdam, The Netherlands.

Kleinbaum, D. G. and Klein, M. (2005). Survival analysis: A self-learning text. 3rd Editio. New York, NY: Springer Science+Business Media.

Kuningas, S., Kvadsheim, P. H., Lam, F.-P. A. and Miller, P. J. O. (2013). Killer whale presence in relation to naval sonar activity and prey abundance in northern Norway. ICES J. Mar. Sci. 70, 1287–1293.

Kvadsheim, P., Lam, F.-P., Miller, P., Doksaeter, 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 Norwegain waters - 3S-2011 cruise report.

Kvadsheim, P. H., Lam, F., Miller, P., Sivle, L. D., Wensveen, P., Roos, M., Tyack, P., Kleivane, L., Visser, F., Curé, C., et al. (2015). The 3S2 experiments - studying the behavioral effects of naval sonar on northern bottlenose whales, humpback whales and minke whales. Horten, Norway.

Lowry, H., Lill, A. and Wong, B. B. M. (2013). Behavioural responses of wildlife to urban environments. Biol. Rev. 88, 537–549.

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.

Maybaum, H. L. (1993). Responses of humpback whales to sonar sounds. J. Acoust. Soc. Am. 94, 1848–1849.

McCarthy, E., Moretti, D., Thomas, L., DiMarzio, N., Morrissey, R., Jarvis, S., Ward, J., Izzi, A. and Dilley, A. (2011). Changes in spatial and temporal distribution and vocal behavior of Blainville's beaked whales (*Mesoplodon densirostris*) during multiship exercises with mid-frequency sonar. Mar. Mamm. Sci. 27, E206–E226.

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.

McFarland, D. J. and Sibly, R. M. (1975). The behavioural final common path. Philos. Trans. R. Soc. L. B 270, 265–293.

Melcón, M. L., Cummins, A. J., Kerosky, S. M., Roche, L. K., Wiggins, S. M. and Hildebrand, J. A. (2012). Blue whales respond to anthropogenic noise. PLoS One 7, e32681.

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

Miller, P. J. 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.

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

Miller, P. J. O., Kvadsheim, P. H., Lam, F. P. A., Tyack, P. L., Curé, C., DeRuiter, S. L., Kleivane, L., Sivle, L. D., van IJsselmuide, S. P., Visser, F., et al. (2015). First indications that northern bottlenose whales are sensitive to behavioural disturbance from anthropogenic noise. R. Soc. Open Sci. 2, 140484.

Mooney, T. A., Yamato, M. and Branstetter, B. K. (2012). Hearing in cetaceans: from natural history to experimental biology. Adv. Mar. Biol. 63, 197–246.

Moretti, D., Thomas, L., Marques, T., Harwood, J., Dilley, A., Neales, B., Shaffer, J., McCarthy, E., New, L., Jarvis, S., 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.

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

New, L. F., Moretti, D. J., Hooker, S. K., Costa, D. P. and Simmons, S. E. (2013a). Using energetic models to investigate the survival and reproduction of beaked whales (family *Ziphiidae*). PLoS One 8, e68725.

New, L. F., Harwood, J., Thomas, L., Donovan, C., Clark, J. S., Thompson, P. M., Cheney, B., Scott-Hayward, L. and Lusseau, D. (2013b). Modelling the

biological significance of behavioural change in coastal bottlenose dolphins in response to disturbance. Funct. Ecol. 27, 314–322.

Nowacek, D. P., Friedlaender, A. S., Halpin, P. N., Hazen, E. L., Johnston, D. W., Read, A. J., Espinasse, B., Zhou, M. and Zhu, Y. (2011). Super-aggregations of krill and humpback whales in Wilhelmina Bay, Antarctic Peninsula. PLoS One 6, e19173.

O'Hara, R. B. and Sillanpää, M. J. (2009). A review of Bayesian variable selection methods: what, how and which. Bayesian Anal. 4, 85–117.

Pater, L. L., Grubb, T. G. and Delaney, D. K. (2009). Recommendations for improved assessment of noise impacts on wildlife. J. Wildl. Manag. 73, 788–795.

Pirotta, E., Merchant, N. D., Thompson, P. M., Barton, T. R. and Lusseau, D. (2015). Quantifying the effect of boat disturbance on bottlenose dolphin foraging activity. Biol. Conserv. 181, 82–89.

Plein, J. P. and Moeschberger, M. L. (2003). Survival Analysis: Techniques for Censoring and Truncated Data. 2nd Edition. New York, NY: Springer.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, pp. 1–10. Vienna, Austria.

R Core Team (2013). R: A language and environment for statistical computing.

Radford, A. N., Purser, J., Bruintjes, R., Voellmy, I. K., Everley, K. A., Wale, M. A., Holles, S. and Simpson, S. D. (2016). Beyond a simple effect: variable and changing responses to anthropogenic noise. In The Effects of Noise on Aquatic Life II, Advances in Experimental Medicine and Biology, pp. 901–907.

Reichmuth, C. (2007). Assessing the hearing capabilities of mysticete whales. A proposed research strategy for the Joint Industry Programme on Sound and Marine Life. http://www.soundandmarinelife.org/libraryfile/1248

Richardson, W. J., Greene, C. R., Malme, C. I. and Thomson, D. H. (1995). Marine Mammals and Noise. San Diego, CA: Academic Press.

Risch, D., Corkeron, P. J., Ellison, W. T. and Parijs, S. M. Van (2012). Changes in humpback whale song occurrence in response to an acoustic source 200 km away. PLoS One 7, e29741.

Schick, R. S., New, L. F., Thomas, L., Costa, D. P., Hindell, M. a, McMahon, C. R., Robinson, P. W., Simmons, S. E., Thums, M., Harwood, J., et al. (2013). Estimating resource acquisition and at-sea body condition of a marine predator. J. Anim. Ecol. 82, 1300–1315.

Sih, A., Bell, A. M., Johnson, J. C. and Ziemba, R. E. (2004). Behavioral syndromes: an integrative overview. Q. Rev. Biol. 79, 241–277.

Simon, R., Freidlin, B., Rubinstein, L., Arbuch, S. G., Collins, J. and Christian, M. C. (1997). Accelarated titration designs for phase I clinical trials in oncology. J. Nat. Can. Inst. 89, 1138–1147.

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.

Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene Jr., C. R., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., et al. (2007). Marine mammal noise exposure criteria: Initial scientific recommendations. Aquat. Mamm. 33, 411–521.

Therneau, T. M. (2015). A Package for Survival Analysis in R.

Tougaard, J., Wright, A. J. and Madsen, P. T. (2015). Cetacean noise criteria revisited in the light of proposed exposure limits for harbour porpoises. Mar. Pollut. Bull. 90, 196–208.

Tyack, P. (1983). Differential response of humpback whales, *Megaptera novaeangliae*, to playback of song or social sounds. Behav. Ecol. Sociobiol. 13, 49–55.

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

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., Samarra, F. I. P., Beerens, S. P. and Miller, P. J. O. (2016). Separating underwater ambient noise from flow noise recorded on stereo acoustic tags attached to marine mammals. J. Exp. Biol. In review.

Wartzok, D., Popper, A. N., Gordon, J. and Merrill, J. (2003). Factors affecting the responses of marine mammals to acoustic disturbance. Mar. Tech. Soc. J 37, 6–15.

Wensveen, P. J. (2012). The effects of sound propagation and avoidance behaviour on naval sonar levels received by cetaceans. MPhil 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. (2015). 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.

Williams, R., Erbe, C., Ashe, E., Beerman, A. and Smith, J. (2014). Severity of killer whale behavioral responses to ship noise: A dose-response study. Mar. Pollut. Bull. 79, 254–260.

Wright, A. J., Maar, M., Mohn, C., Nabe-Nielsen, J., Siebert, U., Jensen, L. F., Baagøe, H. J. and Teilmann, J. (2013). Possible causes of a harbour porpoise mass stranding in Danish waters in 2005. PLoS One 8, e55553.

TABLES AND FIGURES

Table 3.1. Overview of the controlled exposure experiments with humpback whales. An experimental sequence consisted of a *No-Sonar* session followed by two sonar sessions.

Whale ID	**Experiment ID**	**Date and time** (UTC)	**Tagging location** (°N, °E)	**Exposure session**		
				1	2	3
mn11_157a	1	5 June 2011 22:25:15	75.1428, 14.6322	*No-Sonar*	*RampUp1*	*RampUp2*
mn11_158ab	2	7 June 2011 09:21:23	74.8303, 16.6127	*No-Sonar*	-	-
mn11_160ab	3	8 June 2011 22:54:00	74.6108, 15.2939	*No-Sonar*	*RampUp1*	*RampUp2*
mn11_165e	4	14 June 2011 13:40:20	78.0835, 11.0845	*No-Sonar*	*RampUp1*	*RampUp2*
mn11_165df[a)]	4	14 June 2011 13:59:51	78.0840, 11.8113	*No-Sonar*	*RampUp1*	*RampUp2*
mn12_161ab	5	9 June 2012 16:44:27	77.3435, 11.1905	*No-Sonar*	*RampUp1*	*No-RampUp*[b)]
mn12_164ab	6	12 June 2012 17:13:13	77.4770, 09.5942	*No-Sonar*	*RampUp1*	*No-RampUp*[b)]
mn12_170a[a)]	7	18 June 2012 03:31:42	77.6267, 10.4025	*No-Sonar*	*RampUp1*	*RampUp2*
mn12_170b	7	18 June 2012 03:49:48	77.4627, 11.6795	*No-Sonar*	*RampUp1*	*RampUp2*
mn12_171ab	8	19 June 2012 12:21:37	79.0304, 10.6687	*No-Sonar*	*RampUp1*	*RampUp2*
mn12_178a	9	26 June 2012 00:28:20	74.8620, 17.8054	*No-Sonar*	*RampUp1*	*RampUp2*
mn12_179a	10	27 June 2012 07:57:13	74.0454, 20.6807	*No-Sonar*	*RampUp1*	*RampUp2*
mn12_180ab	11	28 June 2012 17:07:52	73.9834, 20.4086	*No-Sonar*	*RampUp1*	*RampUp2*

a) A tagged whale that was associated with the focal whale.
b) Not a dose-escalation experiment. Responses were not used in the dose-response functions.

Table 3.2. Response thresholds of humpback whales for the onset of avoidance and cessation of feeding, quantified by four different dose metrics. Exposure session was coded as "experiment no. – sonar session no." for brevity. Thresholds for sessions in which a response was not observed were censored using the maximum acoustic dose or minimum distance as a boundary.

Exposure session	**Feeding before session**	**SPL_{max}** (dB re 1μPa)			**SEL_{cum}** (dB re 1μPa^2 s)			**S/N** (dB)			**Distance** (m)		
		threshold		max	threshold		max	threshold		max	threshold		min
		avoid	feed		avoid	feed		avoid	feed		avoid	feed	
1-1	N	102		164	97		171	17		63	1389		998
1-2	N			177			181			72			62
3-1	N			174			176			75			129
3-2	N			168			173			72			417
4-1	Y			174			178			72			161
4-2[a)]	N	172		175	174		181	58		69	304		212
5-1	N	132		168	131		174	48		67	1681		858
6-1[b)]	Y	147	125	173	142	119	176	51	35	77	567	891	154
7-1[c)]	Y		164	175		163	180		65	70		585	306
7-2	Y	145		172	141		178	51		75	1632		469
8-1	Y		179	179		179	179		83	83		208	208
8-2	Y		146	182		142	184		48	88		1372	67
9-1	N	143		174	141		180	49		72	1095		499
9-2	Y			174			178			73			232
10-1[a)]	N	135		173	130		179	49		68	1383		348
10-2	N			176			179			70			237
11-1[b)]	Y	165	165	172	160	160	177	58	58	78	795	805	400
11-2	N	127		173	122		177	39		74	995		322

a) Avoidance response was only detectable in fine-scale tracks (see Chapter 2).
b) Onset of avoidance response identified to be slightly later than reported in Sivle et al. (2015).
c) Data of the non-focal whale was used as the focal whale was not feeding.

Table 3.3. Prior probability distributions for all variables in the Bayesian models. Lower and upper limits are reported for uniform distributions (*Unif*) and mean and s.d. are reported for normal distributions (*N*). Parameter values in the model for distance are in units of kilometres. Values for μ in the models for SPL_{max} and SEL_{cum} are in dB re 1 μPa and dB re 1 μPa^2 s, respectively. All other values are in dB. *RampUp1* and non-feeding are the reference levels for session order and behavioural state, respectively. Humpback whale is the reference level for the fixed effects on species being killer whale (β_3) and pilot whale (β_4), respectively.

Variable	**Description**	**SPL_{max}**	**SEL_{cum}**	**S/N**	**Distance**
μ	mean response threshold for all whales	$Unif(60, 200)$	$Unif(60, 200)$	$Unif(0, 120)$	$Unif(0, 2.5)$
φ	between-animal s.d. in response thresholds	$Unif(0, 35)$	$Unif(0, 35)$	$Unif(0, 30)$	$Unif(0, 0.6)$
σ	within-animal s.d. in response thresholds	$Unif(0, 35)$	$Unif(0, 35)$	$Unif(0, 30)$	$Unif(0, 0.6)$
β_1	effect of session order	$N(0, 35)$	$N(0, 35)$	$N(0, 30)$	$N(0, 0.6)$
β_2	effect of behavioural state	$N(0, 35)$	$N(0, 35)$	$N(0, 30)$	$N(0, 0.6)$
β_3	effect of species (1)	$N(0, 35)$	$N(0, 35)$	$N(0, 30)$	$N(0, 0.6)$
β_4	effect of species (2)	$N(0, 35)$	$N(0, 35)$	$N(0, 30)$	$N(0, 0.6)$

Table 3.4. Estimated dose-response model parameters for the onset of avoidance and the cessation of feeding in response to sonar exposure, for models including β-terms. The effect of session order (reference level: sonar session 1) and effect of behavioural state (reference level: not feeding) were quantified by parameters β_1 and β_2, respectively. The "Gibbs Variable Selection *p*-value" indicates the posterior statistical support for including the corresponding fixed effect β. Parameter values in the model for distance are in units of metres. Values for μ in the models for SPL_{max} and SEL_{cum} are in dB re 1 μPa and dB re 1 μPa^2 s, respectively. All other values are in dB.

Dose metric	**Model parameter**	**Onset of avoidance**				**Cessation of feeding**			
		Median	95% CI		GVS *p*	Median	95% CI		GVS *p*
SPL_{max}	μ	128	109	150		177	152	198	
	φ	30	22	35		28	17	35	
	β_1	29	26	31	0.05	–12	–15	–9	0.08
	β_2	65	60	70	0.02				
SEL_{cum}	μ	125	106	147		176	150	198	
	φ	31	23	35		29	19	35	
	β_1	31	29	34	0.05	–10	–14	–7	0.08
	β_2	67	62	72	0.02				
S/N	μ	42	27	58		73	54	99	
	φ	23	15	30		24	14	30	
	β_1	19	16	21	0.06	–12	–15	–9	0.09
	β_2	38	33	42	0.07				
Distance	μ	988	608	1318		297	18	707	
	φ	512	362	596		458	271	592	
	β_1	382	356	409	0.03	–374	–408	–338	[a]
	β_2	585	537	636	0.05				

a) Model did not run with tuned GVS pseudo-prior

Table 3.5. Estimated dose-response model parameters for the onset of avoidance and the cessation of feeding in response to sonar exposure, for models excluding the β-terms but including within-animal variation in thresholds (σ). The estimated dose at which 50% of the humpback whales was predicted to respond (p_{50}) is also given. Parameter values in the model for distance are in units of metres. Values for μ in the models for SPL_{max} and SEL_{cum} are in dB re 1 μPa and dB re 1 μPa^2 s, respectively. All other values are in dB.

Dose metric	**Model parameter**	**Onset of avoidance**			**Cessation of feeding**		
		Median	95% CI		Median	95% CI	
SPL_{max}	μ	180	156	199	178	151	199
	φ	16	1	33	18	1	34
	σ	31	23	35	26	12	34
	p_{50}	166	151	176	166	150	175
SEL_{cum}	μ	178	151	199	178	149	199
	φ	18	1	34	19	1	34
	σ	26	12	34	27	14	35
	p_{50}	168	148	179	167	146	178
S/N	μ	72	55	97	73	51	106
	φ	13	1	29	14	1	29
	σ	26	18	30	23	12	30
	p_{50}	71	56	86	71	52	92
Distance	μ	300	17	724	306	16	807
	φ	323	28	582	264	14	574
	σ	560	451	598	497	317	595
	p_{50}	578	427	842	528	352	879

Table 3.6. Estimated dose-response model parameters for the onset of avoidance in response to sonar exposure for the models with 3 cetacean species shown in Figure 3.12. SPL_{max} was used as the dose metric. Independent models were fitted to all avoidance thresholds (1-2 kHz and 6-7 kHz sonar) and to avoidance thresholds observed during 1-2 kHz sonar sessions only, as humpbacks were not exposed to 6-7 kHz sonar. The bottom rows list the estimates of the received SPLs at which 50% of the humpback whales (*hw*), killer whales (*kw*) or long-finned pilot whales (*pw*) were predicted to have started to respond (p_{50}).

Parameter	3-Species model for LFAS and MFAS				3-Species model for LFAS only			
	Median	95% CI		GVS *p*	Median	95% CI		GVS *p*
μ	180	158	199		180	158	199	
φ	19	1	34		16	1	33	
σ	32	24	35		32	24	35	
β_3	−27	−51	2	0.31	−16	−48	20	0.40
β_4	17	−9	45	0.36	15	−15	48	0.37
$p_{50,hw}$	166	153	175		167	154	175	
$p_{50,kw}$	149	127	166		157	130	175	
$p_{50,pw}$	173	160	182		174	157	184	

Figure 3.1. Selection of a segment of background noise (blue dashed lines) used for the calculation of signal-to-noise ratio (S/N). The panels show the: (A) recorded acoustic pressure in the 1.3-2 kHz frequency band, and source level and duration of three transmitted sonar pulses; (B) spectrogram of the same recording, where the colour gradient indicates the power spectral density level (in dB re 1 μPa^2 Hz^{-1}); (C) sound pressure level (SPL) in the 1.3-2 kHz band, with markers indicating in which time bins flow noise (red) or other noise (green) was the dominant contribution to the SPL; and (D) the depth of the whale and its speed through the water. In this example, the SPL_{max} was calculated from the sonar pulse starting at 20 s, but its S/N was based on the noise level of the subsequent sonar pulse due to the presence of flow noise.

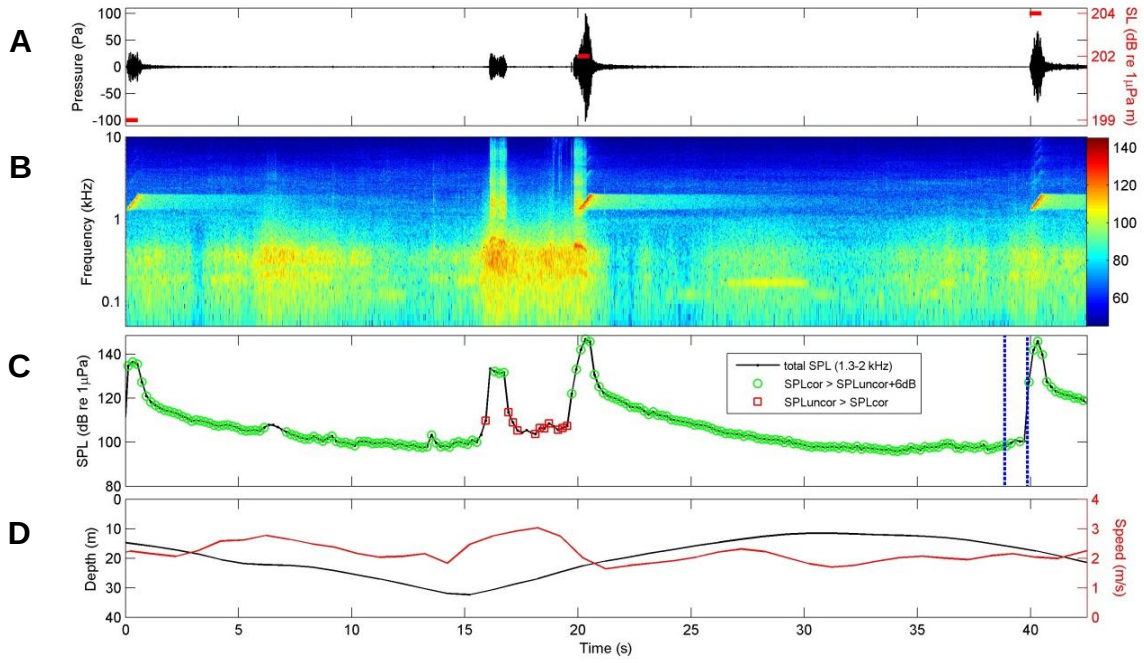


Figure 3.2. Selection of a segment of background noise (blue dashed lines) used for the calculation of signal-to-noise ratio (S/N). See the legend of Figure 3.1 for a full description of the data series that are plotted. The automatic method to separate flow noise from other noise (von Benda-Beckmann et al., 2016) was not used in this case because the DTAG had only one hydrophone. Visual inspection was used instead to ensure that the selected noise levels were not significantly affected by flow noise. In this example, the segment was dominated by reverberation caused by the previous pulse, and the S/N used was thus a signal-to-reverberation ratio.

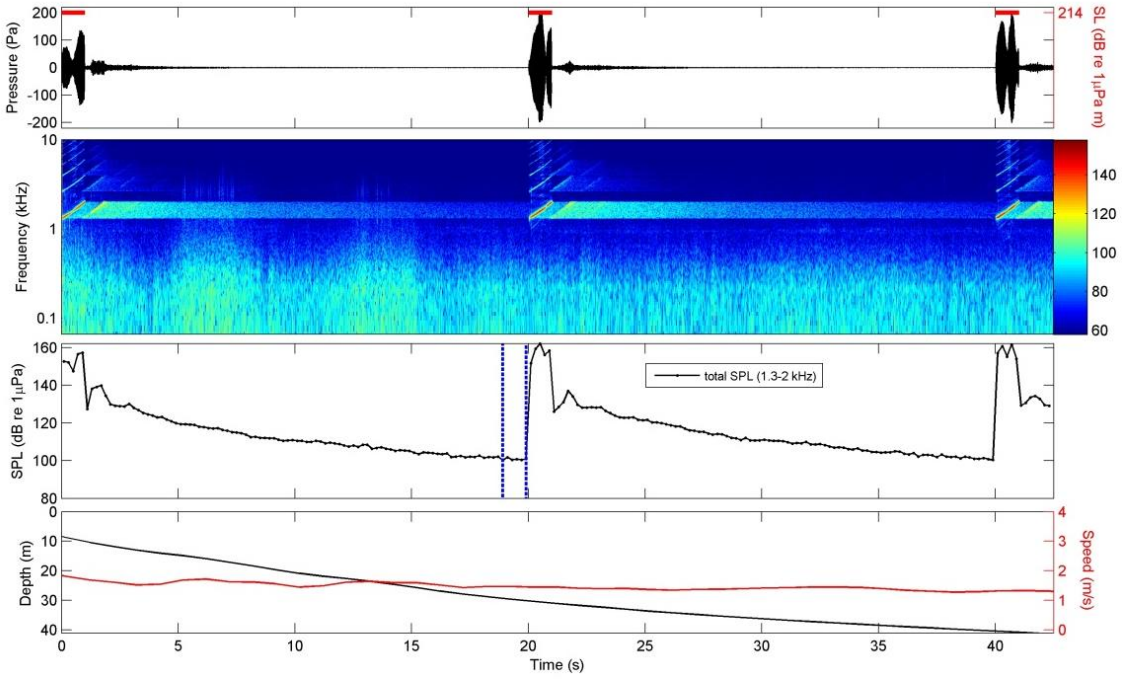


Figure 3.3. Selection of a segment of background noise (blue dashed lines) used for the calculation of signal-to-noise ratio (S/N). See the legend of Figure 3.1 for a full description of the data series that are plotted. In this example, the relative contribution of flow noise was low but visual and aural inspection of the recording showed that vessel noise dominated the background noise just prior to the pulse.

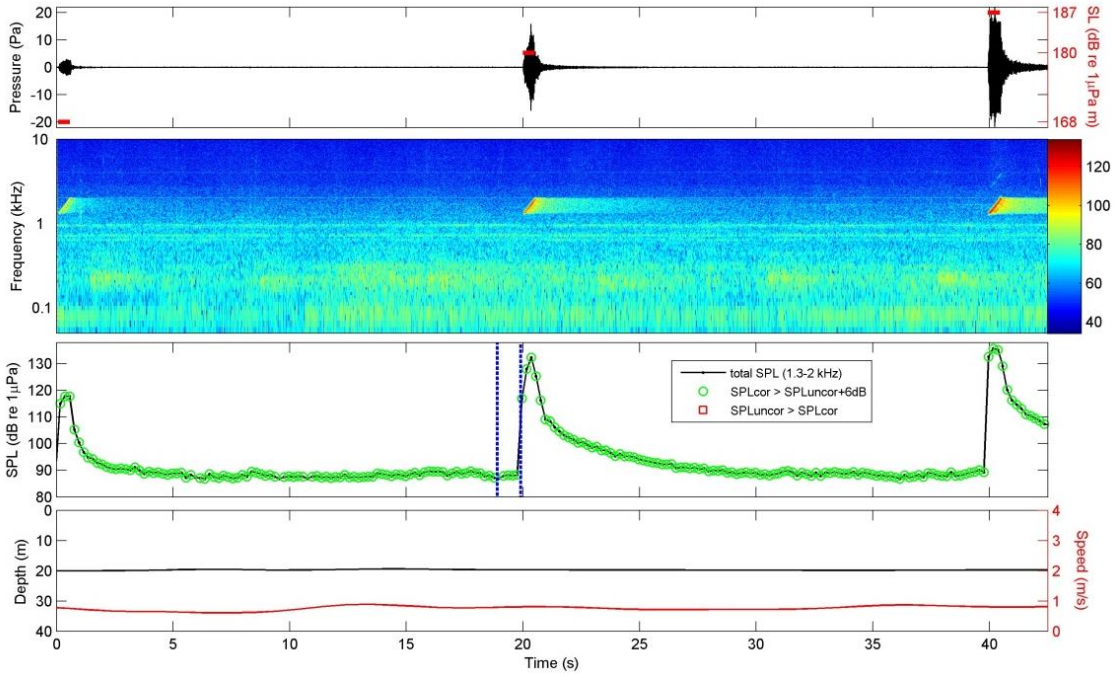


Figure 3.4. Prior dose response functions for all dose metrics. For each panel, the median is represented by the solid line and the 95% credible intervals (CIs) are indicated by the dashed lines.

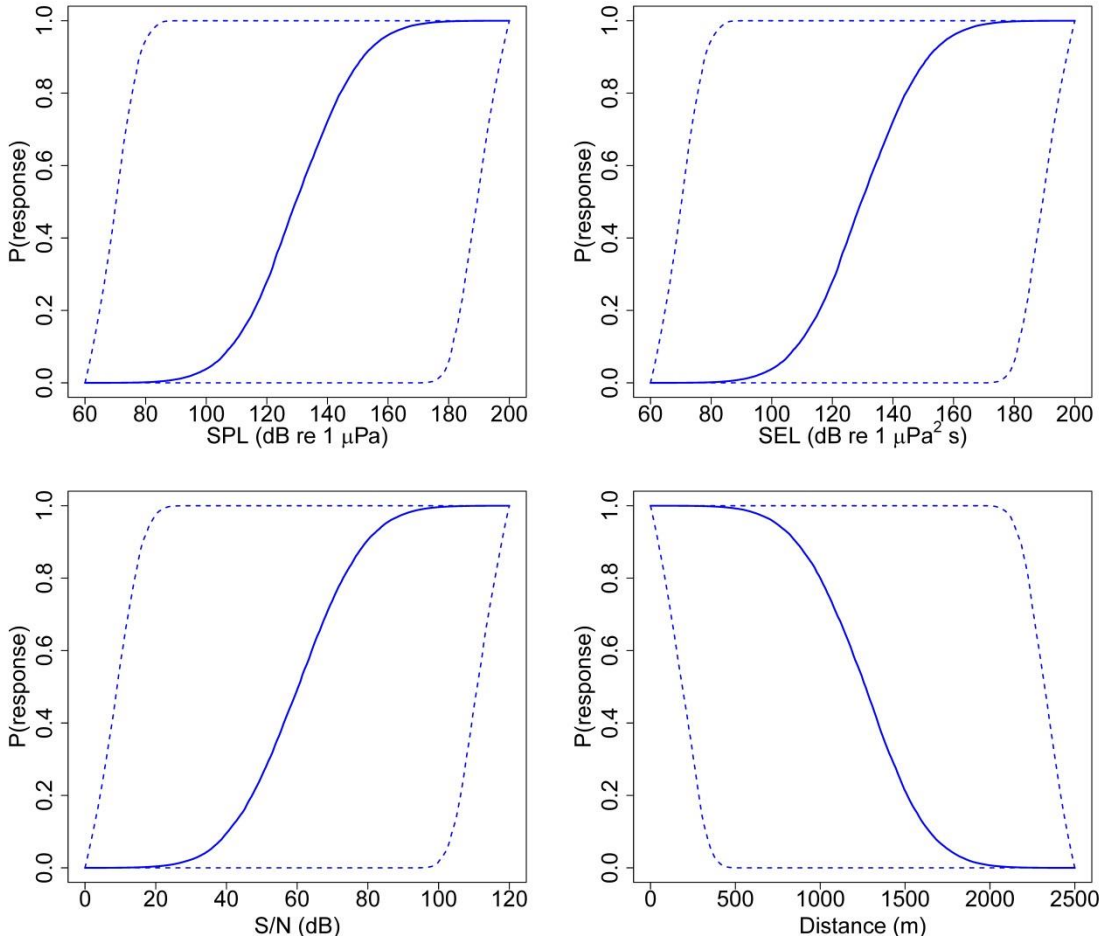


Figure 3.5. (A-C) Correlations between the different dose metrics and (D) the correlation between the SPL_{max} of the sonar signal and the ambient noise level when the signal was received. Markers indicate the observed response thresholds (*y*) of the whales. Pearson's correlation coefficient *r* is shown in each panel.

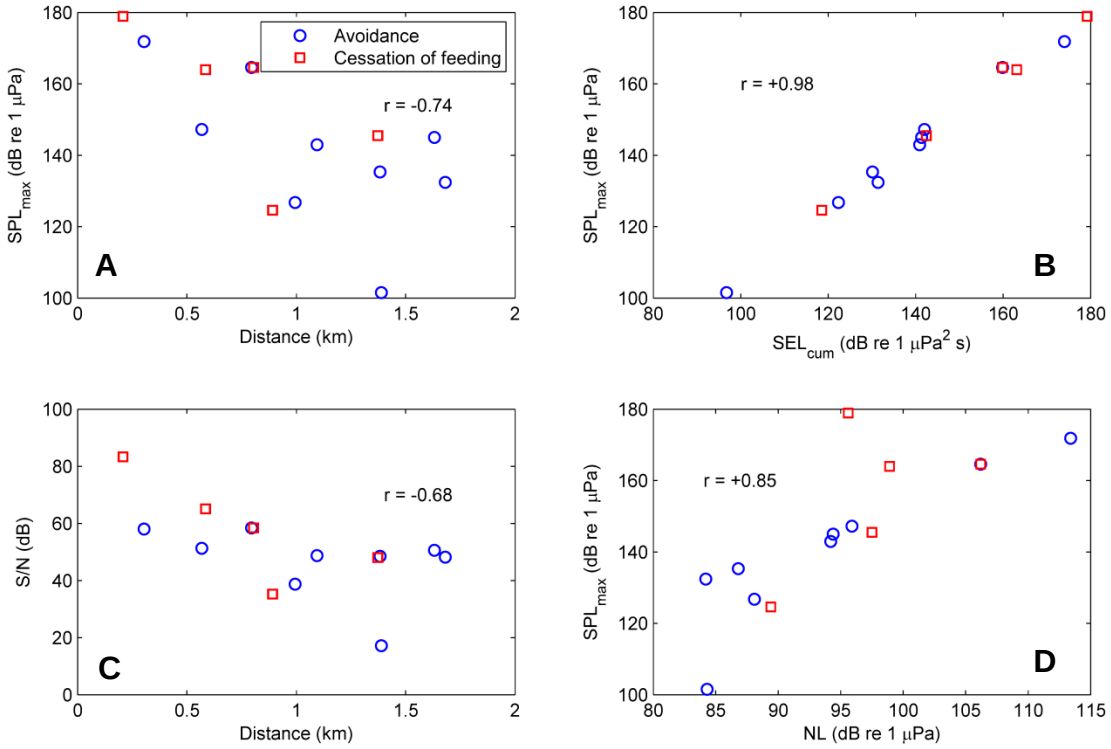


Figure 3.6. Doses of sound or source distance to which the whales were exposed. Grey lines show the range where a response was not identified and black lines show the range beyond the observed response. Blue circles and red squares correspond to the observed response thresholds (*y*) for onset of avoidance and cessation of feeding, respectively.

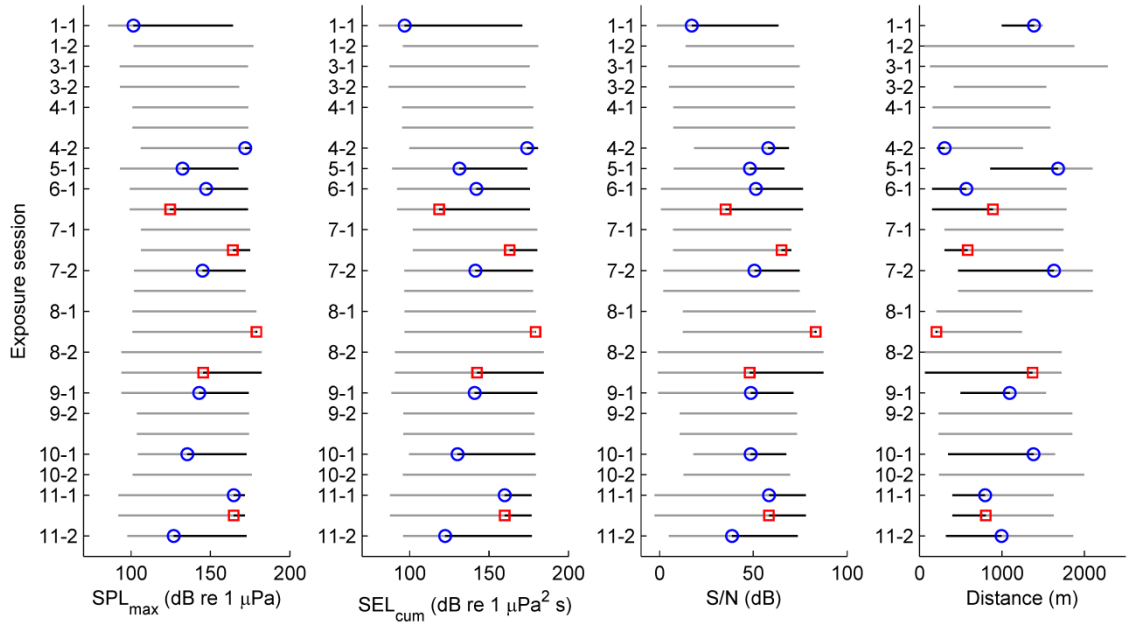


Figure 3.7. (columns on the left) Kernel smoothed density estimates of the prior (grey) and posterior (black) probability distributions for the estimated parameters in the full SPL_{max} model for onset of avoidance. (columns on the right) Trace plots of the two MCMC chains (black and red) for the same model. The gamma parameters indicate the level of support in the data for including the corresponding beta terms (1 or 2), which were used to obtain the "GVS *p*-value", i.e. the proportion of posterior samples in which the parameter takes the value 1. The prior on each gamma parameter was a Bernoulli distribution with success probability 0.5.

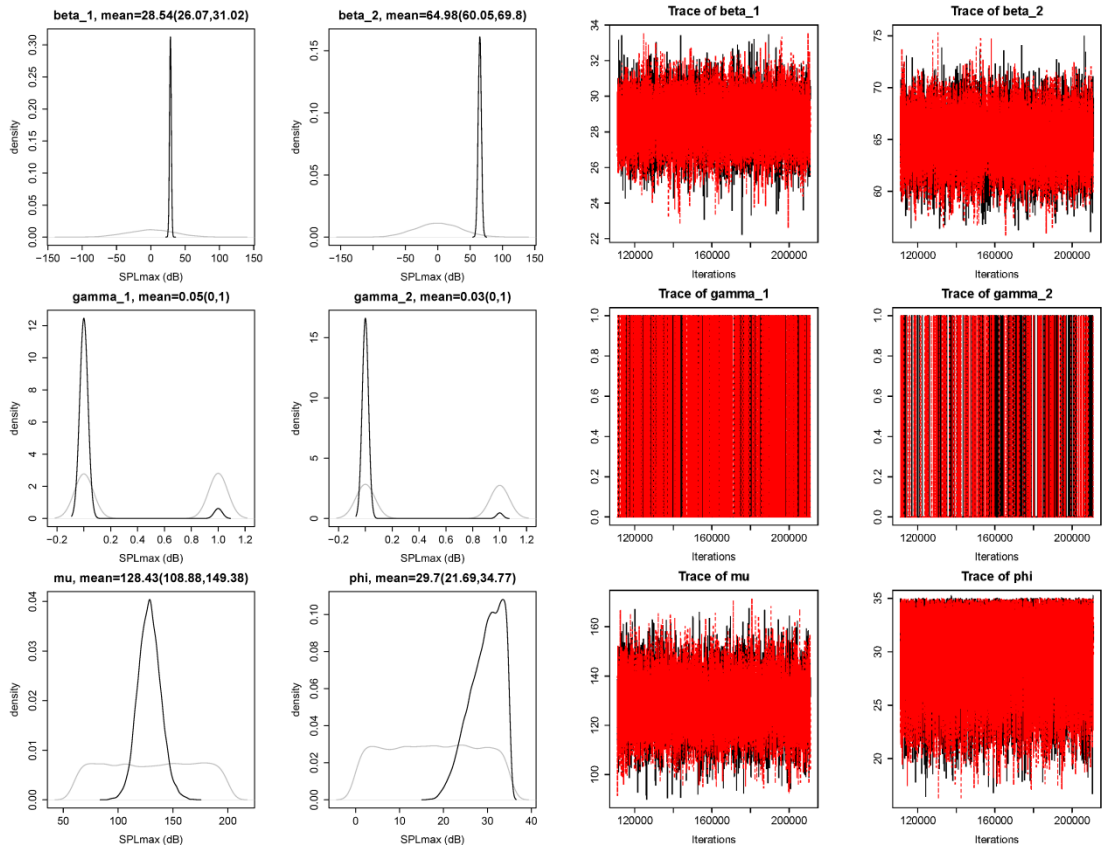


Figure 3.8. The evolution of the Brooks-Gelman-Rubin statistic $\hat{R}$ (or, "shrink factor") as the number of MCMC iterations increased ($\hat{R}$ values of <1.05 indicate convergence of the chains) for the same model as in Fig. 3.7. Similarly quick convergence was obtained for the other models.

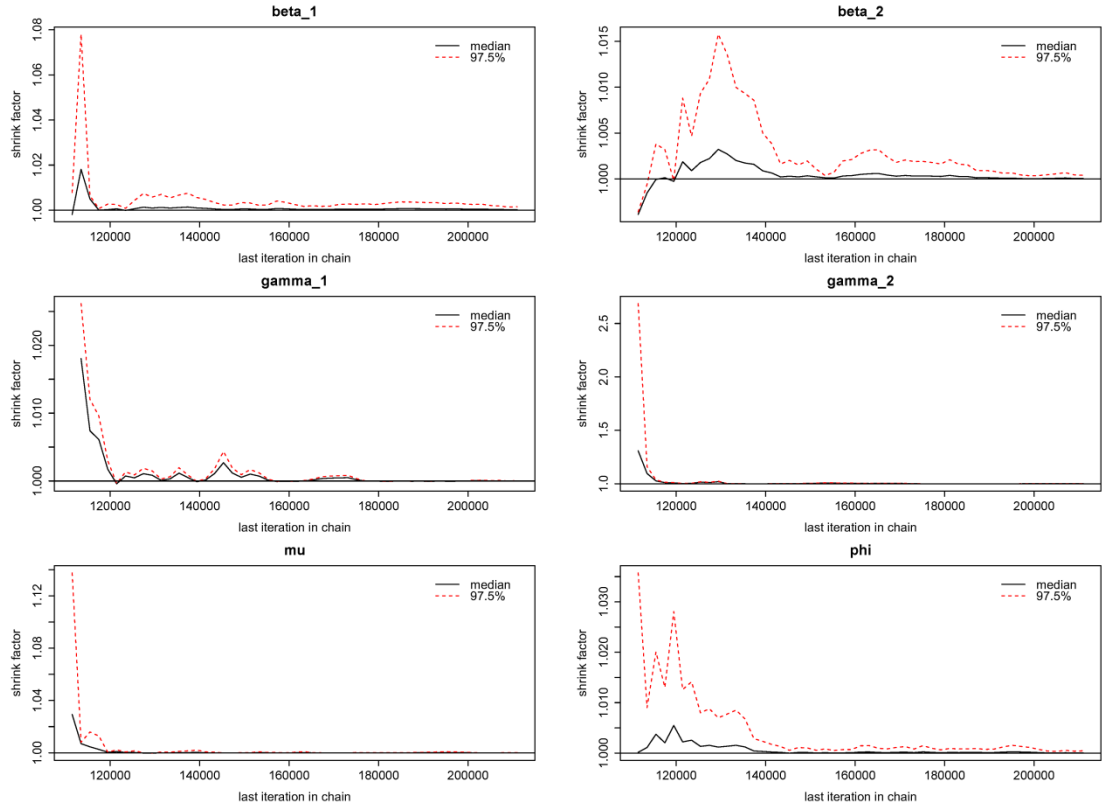


Figure 3.9. Context-specific dose-response curves for the onset of avoidance by humpback whales in response to sonar exposure. The four panels correspond to animals in a (A) non-feeding state during sonar session 1, (B) non-feeding state during sonar session 2, (C) feeding state during sonar session 1 and (D) feeding state during sonar session 2. Solid lines are median posterior dose-response curves and dashed lines indicate their 95% CIs. The posterior median and 95% CI of the estimated p_{50} (i.e. the dose at which the model predicts that 50% of animals have started to avoid) is shown in the top left corners. The raw data are shown in the bottom of the panel (circles: observed response thresholds; triangles: maximum doses of the sessions in which a response was not observed).

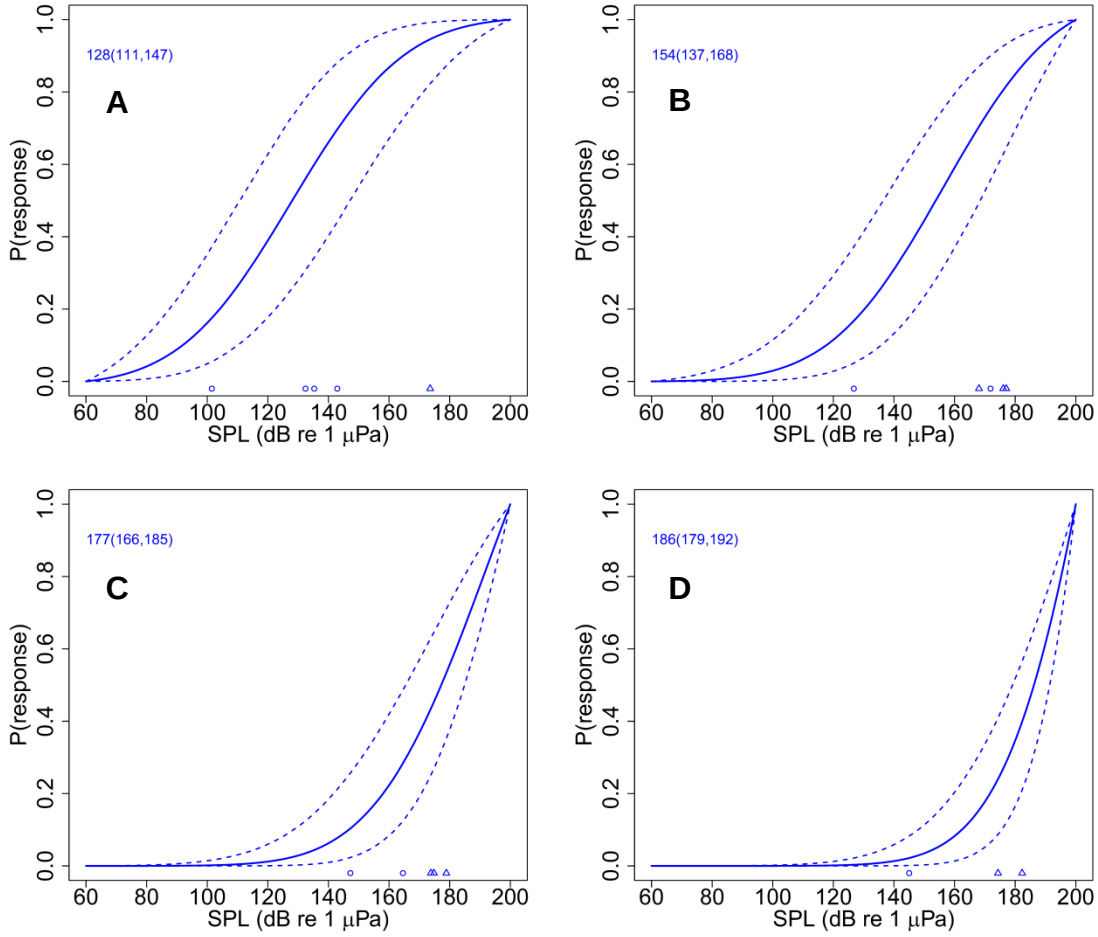


Figure 3.10. Dose-response curves for exposures of 1.3-2 kHz sonar signals to humpback whales for the dose metrics analysed, calculated from the null models (i.e. the models without beta terms but including a random effect for within-animal variation). Blue and red curves correspond to the onset of avoidance and cessation of feeding, respectively. Solid lines are medians and dashed lines indicate 95% CIs. The raw data are shown in the bottom of the panel (circles: observed response thresholds; triangles: maximum dose of the sessions in which a response was not observed).

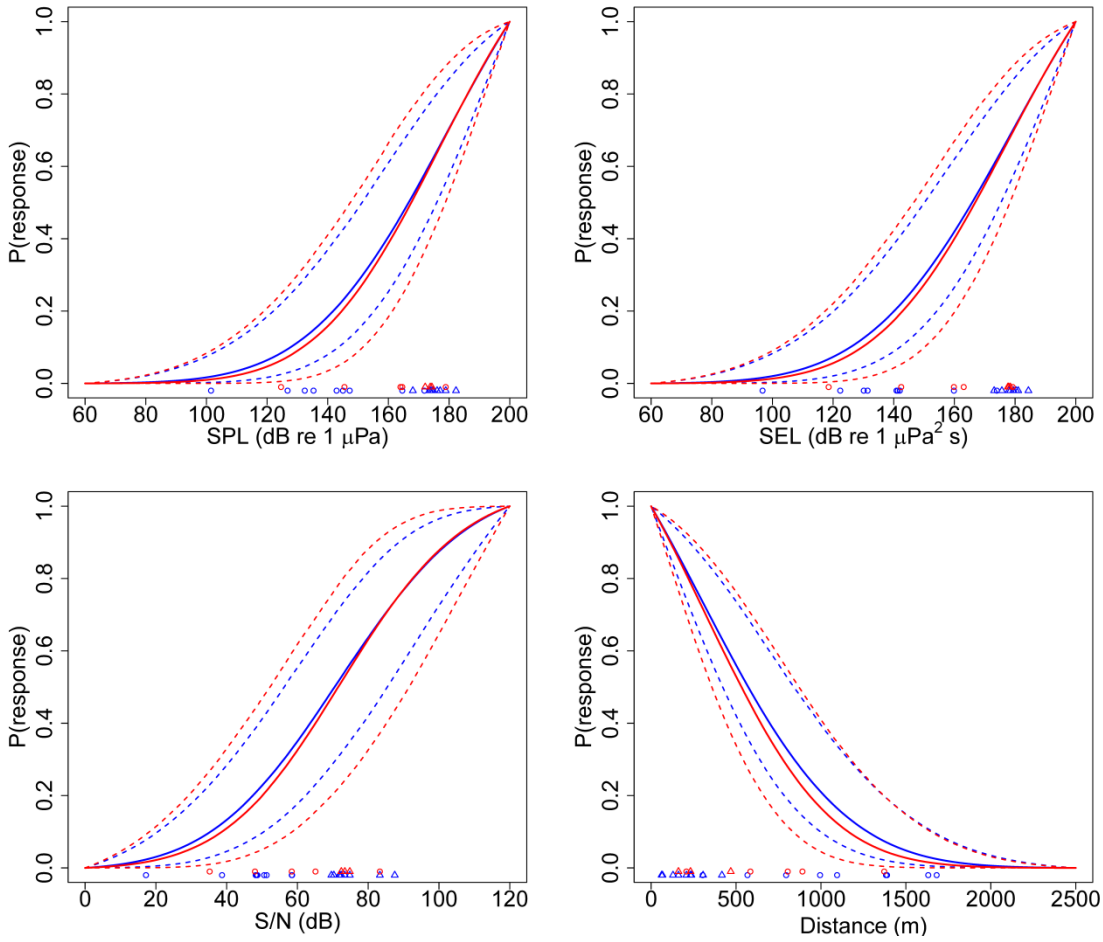


Figure 3.11. Dose-response curves for humpback whales exposed to naval sonar, obtained using recurrent event survival analysis (Harris et al., 2015; black; all response types) and the Bayesian model detailed in 'Methods - Fitting the dose-response functions' (blue: onset of avoidance; red: cessation of feeding). Dashed lines indicate 95% CIs. Both methods were applied to data from the same dose-escalation experiments, but the data set used in the survival analysis contained thresholds for additional types of response such as changes in dive behaviour and changes in group size. The black curve is valid for response severities 4 to 6 (Southall et al., 2007); all thresholds underlying the Bayesian model fell within this category (Sivle et al., 2015).

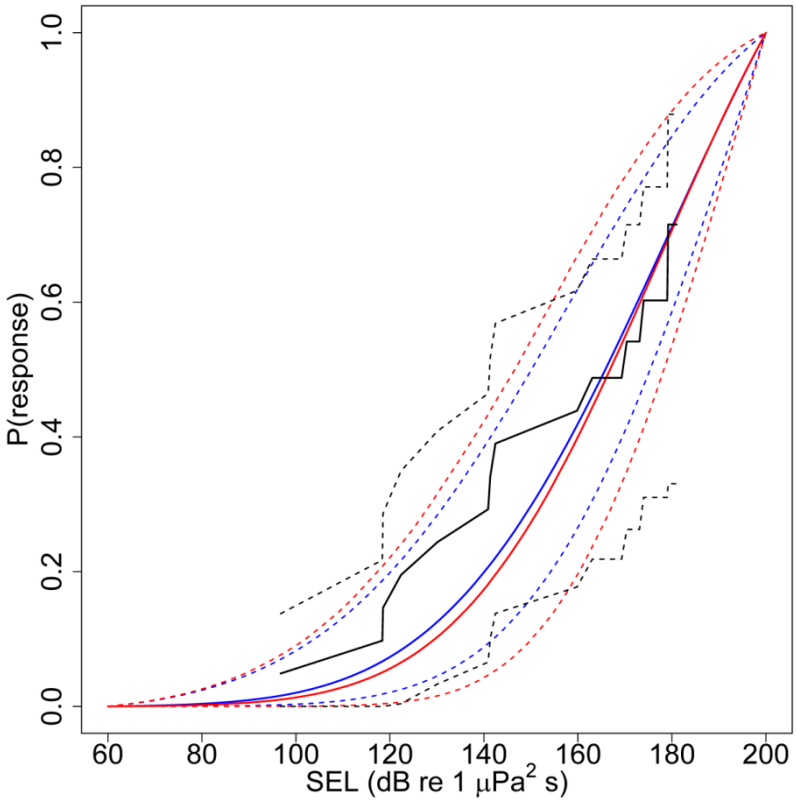
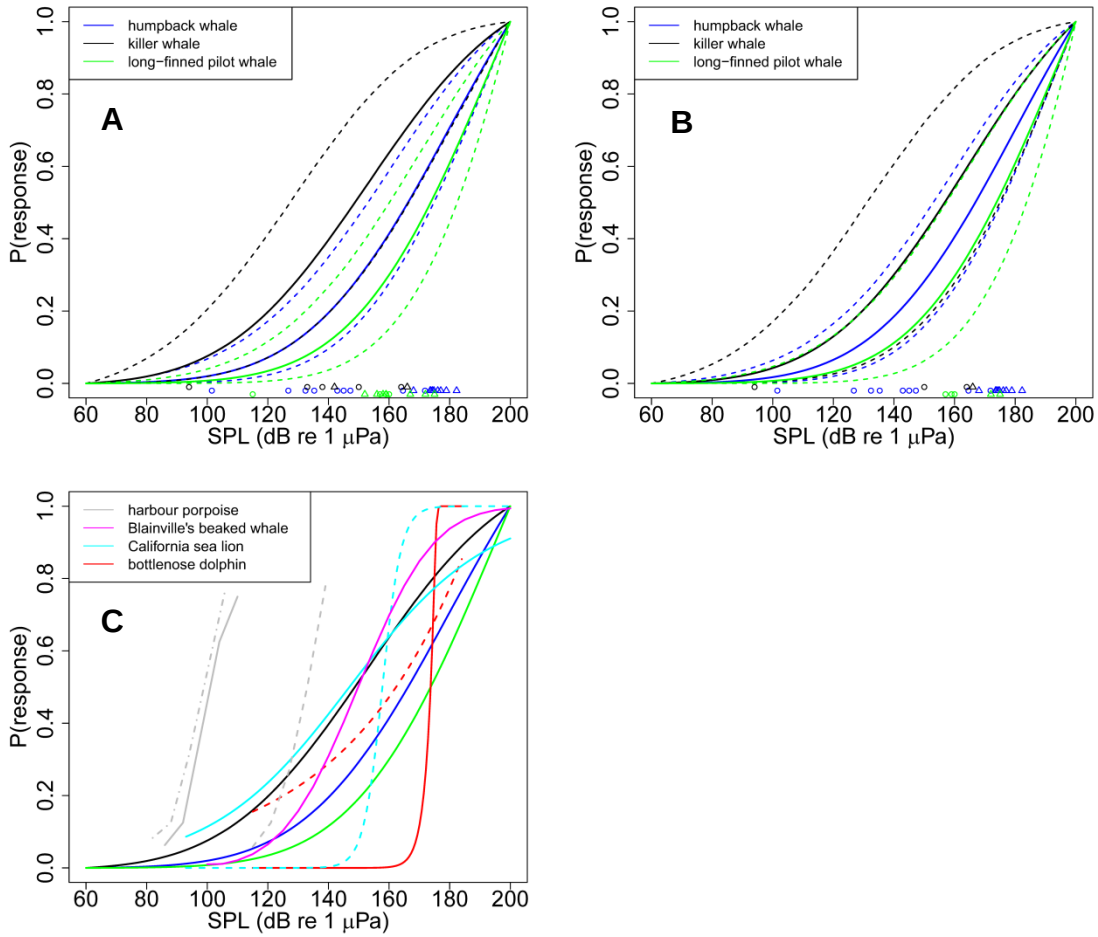


Figure 3.12. Median posterior dose-response curves and 95% CIs for avoidance responses of humpback (this chapter), killer whales (Miller et al., 2014) and long-finned pilot whales (Antunes et al., 2014) in response to exposures of (A) both 1-2 and 6-7 kHz signals and (B) 1-2 kHz signals only. (C) Visual comparison of the curves shown in panel A with the published curves from an observational study: Moretti et al. (2014) for cessation of foraging clicks in response to predominantly 3-4.5 kHz signals in Blainville's beaked whales, and from three laboratory studies: Kastelein et al. (2012) for brief changes in orientation and speed in a harbour porpoise (solid line: 6-7 kHz sweeps, dashed line: 1-2 kHz sweeps, dashed-dotted line: 1-2 kHz sweeps with harmonics) and Houser et al. (2013a, 2013b) for various types of behavioural response to 3.3-3.5 kHz signals in California sea lions (solid line: 1-29 years of age, dashed line: 3-29 years of age) and bottlenose dolphins (solid line: trial #1, dashed line: trial #10).



Chapter 4

Auditory weighting functions based on behavioural reaction times of a harbour porpoise

SUMMARY

Loudness perception by human infants and animals can be studied under the assumption that sounds of equal loudness elicit equal behavioural reaction times (RTs). Therefore, frequency weighting functions based upon equal-reaction time contours may be useful for improving noise exposure criteria for marine mammals. To test the feasibility of this approach, simple RTs of a trained harbour porpoise to narrowband tonal signals were measured using a behavioural method and an RT sensor based on infrared light. Equal latency contours, which connect equal RTs across signal frequencies, for reference values of 150, 160, 170, 180, 190 and 200 ms were derived from median RTs to 1-s signals with sound pressure levels (SPLs) of 59–168 dB re 1 μPa and centre frequencies of 0.5, 1, 2, 4, 16, 31.5, 63, 80 and 125 kHz. The equal latency contours roughly paralleled the hearing threshold at relatively low sensation levels (i.e. higher RTs). Differences in shape between the audiogram and equal latency contours were more pronounced at higher levels (i.e. lower RTs); a flattening of the contours occurred for frequencies below 63 kHz. Relationships of the equal latency contour levels with the audiogram were used to create smoothed functions assumed to be representative of equal loudness contours and auditory weighting functions were derived from these smoothed functions. A function with a relatively steep low frequency roll-off (45-49 dB/decade) agreed most closely with behavioural response onset thresholds for captive and wild harbour porpoises, while the limited data on onset of temporary threshold shift (TTS) that are currently available for harbour porpoises tentatively suggested better correspondence to a contour with a shallower low frequency roll-off (32-37 dB/decade). Finally, the weighting functions were compared to published information on the sound levels needed to elicit TTS and behavioural responses in captive and wild harbour porpoises.

INTRODUCTION

Sound exposure guidelines for anthropogenic noise exposure to marine mammals often contain some form of frequency-selective weighting according to the perception of the target species, so that single acoustic risk thresholds apply to many sounds irrespective of their frequency spectra. Weighted risk thresholds in sound exposure guidelines have been obtained using auditory weighting functions based on the audiogram (e.g. Nedwell et al., 2006; Verboom and Kastelein, 2005) or the approximate frequency bandwidth of hearing (M-weighting) (Southall et al., 2007), but these two methods often produce very different weighted levels (e.g. De Jong and Ainslie, 2008). For humans, weighted thresholds are generally obtained using weighting functions derived from equal loudness contours, e.g. the A- and C-weighting functions (Kinsler et al., 2000).

Equal loudness contours present the relationship between the received sound pressure level (SPL) and the perceived loudness across frequency (Fletcher and Munson, 1933; Suzuki and Takeshima, 2004). Finneran and Schlundt (2011) directly measured the equal loudness contours of a bottlenose dolphin (*Tursiops truncatus)*. The dolphin was presented with a test tone and a reference tone in each trial, and was trained to indicate whether the test tone was louder or softer than the reference tone. It was difficult to convey the complex task to the dolphin, and thousands of trials had to be completed before the equal loudness contours were obtained. The three equal loudness contours were comparable in shape to the animal's audiogram, and became somewhat shallower as loudness increased, as expected from human equal loudness contours. An auditory weighting function derived from one of the contours closely agreed with the temporary threshold shift (TTS) onset thresholds of two bottlenose dolphins (Finneran and Schlundt, 2013).

Perceived loudness is a subjective descriptor of sound that is difficult to quantify in animals. It is more practical to measure simple reaction time (RT; or response latency) to a sound, which correlates with loudness (for review: Luce, 1986; Marks and Florentine, 2011). Simple RT is defined as the time that elapses between the onset of a stimulus and the initiation of a response, when only one type of response is possible. In

humans, a strong correlation between RT and perceived loudness has been demonstrated by loudness comparison tests with pure tones (Buus et al., 1982; Kohfeld et al., 1981) and 1/3-octave bands (Humes and Ahlstrom, 1984), and by exploiting temporal and spectral loudness effects, such as loudness recalibration (Arieh and Marks, 2003), softness imperfection (Florentine et al., 2004), and spectral summation of loudness (Wagner et al., 2004). Equal latency contours, which describe the frequency-dependent relationships between SPL and RT, are similar in shape to equal loudness contours in humans (Marshall and Brandt, 1980; Pfingst et al., 1975a).

In animals, equal latency contours have been obtained for the crab-eating macaque (*Macaca irus*) (Stebbins, 1966), common squirrel monkey (*Saimiri sciureus*) (Green, 1975), rhesus macaque (*Macaca mulatta*) (Pfingst et al., 1975a, 1975b), house finch (*Carpodacus mexicanus*) (Dooling et al., 1978), and domestic cat (*Felis catus*) (May et al., 2009); preliminary equal latency data have been obtained for the harbour seal (*Phoca vitulina*) (Kastelein et al., 2011b) and California sea lion (*Zalophus californianus*) and bottlenose dolphin (*Tursiops truncatus*) (Mulsow and Finneran, 2013; Ridgway and Carder, 2000). The equal latency contours of the animals tested to date are similar to the equal loudness contours of humans and the bottlenose dolphin, which suggests that RTs are also related to perceived loudness in other animals. Hence, frequency weighting based on equal latency may be a relatively time-efficient alternative to frequency weighting based on equal loudness, for all marine mammal species that can be trained with behavioural methods.

In this thesis chapter, underwater equal latency contours were measured in a harbour porpoise, *Phocoena phocoena* responding behaviourally to narrowband frequency-modulated (FM) sound signals with a wide range of centre frequencies and SPLs. Based on the results, relationships between the equal latency contours and the audiogram of the porpoise were determined to create smoothed functions that are assumed to be representative of the equal loudness contours of the animal. The smoothed functions were then used to derive a family of auditory weighting functions for the harbour porpoise that can be used to predict perceived levels and correlated effects of noise. Finally, the weighting functions were compared to published information on the sound

levels needed to elicit TTS and behavioural responses in captive and wild harbour porpoises.

METHODS

Test facility and study subject

The subject was a male harbour porpoise (Jerry; ID 02) that had been rehabilitated after being stranded at the age of about 21 months. The porpoise was well trained and had participated in a number of psychoacoustic studies, including recent studies on temporary threshold shift (Kastelein et al., 2012a, 2014, 2015b). Veterinary records of the animal showed no exposure to ototoxic medication. The porpoise's body condition (body mass, length, girth and blubber thickness) was checked once a week to ensure that he was healthy and at his target body mass. This study was conducted in 2011 and 2012, during which the animal aged from 6 to 7 years, weighed 39 kg, his body length was 145 cm, and his girth at axilla was 73 cm. The animal received about 2 kg of thawed fish per day and was fed four times a day, in general during research sessions.

Test sessions were conducted at the SEAMARCO Research Institute, the Netherlands; a facility for psychophysical research located in a remote and quiet area. The test sessions were performed in an indoor test pool (8 m × 7 m, 2 m deep; Fig. 4.1) that was part of the porpoise's own pool complex. To absorb sound energy from reflections, the walls were covered with 3 cm thick coconut mats with their fibres embedded in 4 mm thick rubber (most effective at >25 kHz), and the bottom of the pool was covered with a 20 cm thick layer of sloping sand on which aquatic vegetation grew.

The water temperature during the study varied between 14 and 18°C, and the salinity was around 34‰. The water pumps and air pumps for the research pool and neighbouring pools were shut off 15 minutes before test sessions commenced. By the time a session had started, little to no water flowed over the skimmers and through the pipes, reducing the influence of flow noise on the background noise level. Information on the water circulation and aeration systems can be found elsewhere (Kastelein et al., 2009).

To avoid distracting the animal, nobody was allowed to move within 15 m of the research pool during sessions. The signal operator and the equipment used to produce the sound signals were out of sight of the animal at the listening station, in a research cabin next to the indoor pool (Fig. 4.1). The listening station was at the end of a 32-mm diameter water-filled polyvinyl chloride tube, 1 m below the water surface (i.e. mid-water).

Sound stimuli

The sound stimuli were narrowband sinusoidal FM signals with centre frequencies of 0.5, 1, 2, 4, 16, 31.5, 63, 80 and 125 kHz. The signals were created digitally in MATLAB (version 7.5; The Mathworks, Natick, MA, USA) using the FM synthesis equation (Chowning, 1973). The frequency deviation was 2% of the centre frequency, and the modulation frequency was 100 Hz. Therefore, for example, when the centre frequency was 1 kHz, the actual frequency of the signal fluctuated 100 times per second between 0.99 and 1.01 kHz. FM stimuli were used because in small, reverberant pools such signals produce a more uniform sound field with fewer standing waves than pure tones (Finneran and Schlundt, 2007). The duration of the test signal was always 1 s. Each signal was cosine-tapered to create a 50 ms ramp on either side of the waveform (10% Tukey window), in order to prevent onset and offset clicks and reduce the probability of eliciting startle reflexes (rise time is positively related to startle reflex thresholds in mammals; Fleshler, 1965; Götz and Janik, 2011).

The sound pressure level (SPL) of the stimuli received by the porpoise while at the listening station ranged from 59 to 168 dB re 1 μPa (depending on the frequency), and test levels were spaced 10 dB apart. The minimum test level varied across frequencies from 3 to 22 dB in terms of sensation level (SnL; *sensu* Ellison et al., 2012). Sensation level is defined here as the number of dBs above the subject's 50% detection hearing threshold for 900 ms tonal signals, which was measured 2-3 years earlier by Kastelein et al. (2010) in this harbour porpoise. The maximum test levels were determined *a priori* based on two criteria: 1) signals could not induce hearing threshold shift in the animal or cause adverse behavioural responses (e.g. hesitation to approach the listening station

after a trial with a high level), and 2) the SPL of any given harmonic had to be at least 30 dB below the SPL of the fundamental frequency.

Acoustic equipment

The equipment used to generate and transmit the sounds, record the electrical signals from the reaction time sensor (see 'Reaction time measurements', below), and monitor the animal's behaviour and the underwater sound field, is shown in Fig. 4.2. The digital sound signals (sample rate: 1 MHz) were converted to analog signals using a 16-bit data acquisition (DAQ) device (National Instruments USB-6251 BNC, Austin, TX, USA) connected to a laptop computer. To increase the dynamic range of the transmission system, the electric output of the DAQ card went through a custom-built digitally-controlled attenuator (AS 2009-01, Smink, Harderwijk, the Netherlands) before going to the projector. The attenuator also functioned as a low-pass reconstruction filter.

Four projectors were used to transmit the signals into the water (Fig. 4.2). The 0.5-2 kHz signals were first fed into an audio power amplifier (Vellerman HQ VPA2450MB, Gent, Belgium) and then transmitted by a high-power piezoelectric projector [Lubell Labs (LL) 1424HP, Columbus, OH, USA] driven by an isolation transformer (LL AC1424HP). This projector was also used to transmit 4-kHz signals of SnLs $\geq$ 48 dB, but in other sessions, only 4-kHz signals of SnLs $\leq$ 58 dB were transmitted unamplified and with a balanced tonpilz piezoelectric projector (LL 916) driven by an isolation transformer (LL AC202). The 16-63 kHz signals were transmitted by a cylindrical piezoelectric projector (International Transducer Corporation 6084, Santa Barbara, CA, USA). The 80-125 kHz signals were transmitted by a custom-built discoid piezoelectric projector (WAU q7b, Honolulu, HI, USA; for more details, see Kastelein et al., 2009).

To minimize temporal and spatial variations in the underwater sound field caused by multi-path arrivals, all projectors except the LL 1424HP were placed in a corner of the pool in a protective wooden box (Fig. 4.1), which was lined with rubber with an irregular surface. These projectors were 2 m from the porpoise's external auditory meatus while the animal was at the listening station. The high-power LL 1424HP did not fit in the protective box, so this projector was hung in front of the box by ropes

attached to its stainless steel cage, at 1.2 m from the porpoise's external auditory meatus (Fig. 4.1). The directional WAU q7b projector was positioned so that the acoustic beam axis pointed at the centre of the porpoise's head. A baffle board with a 30-cm diameter hole was placed halfway between the projector and the animal to reduce reflections from the bottom of the pool and the water surface reaching the listening station. The board was made of 2.4 m high, 1.2 m wide, 4 cm thick plywood, covered with a 2 cm thick closed-cell rubber mat on the side facing the projector.

The output of the sound system was checked before every session with a digital storage oscilloscope (Voltcraft 632FG, Hirschau, Germany) and a voltmeter (Hewlett Packard 3478A, Palo Alto, CA, USA), by playing a signal with a known root-mean square (RMS) voltage from the computer. The test signals and background noise in the water were monitored using the same oscilloscope and voltmeter. Before and during sessions, the system was further verified by listening to the underwater sound via a monitoring hydrophone (Labforce 1 90.02.01, Gouda, the Netherlands) positioned next to the hole in the baffle board. The output of the monitoring hydrophone was fed into either a charge amplifier [Bruel & Kjær (B&K) 2635, Nærum, Denmark] and amplified loudspeaker, or a modified ultrasound detector (Batbox III, Steyning, UK).

Response measurement system

An optical sensor system to measure the animal's responses was designed and built for this study. The reaction time sensor's electronic circuit consisted of an infrared detector integrated circuit (Sharp IS471FE, Osaka, Japan) connected to a 319 THz narrow-beam infrared light emitting diode (LED). The intensity of the infrared light was modulated (38 kHz frequency) by the integrated circuit, making the detector impervious to disturbing external light. The electronic components were embedded in transparent polyurethane epoxy, inside two bracket-shaped polyvinyl chloride pipes (see Figs. 4.1B and 4.3). The infrared emitter and detector were placed directly above and below the tip of the listening station, respectively, spaced 13 cm apart, and facing each other. The tip of the listening station reached just inside the effective optical beam, which was about 8 mm in diameter at that location. The sensor indicated 'presence' when the infrared light was blocked by the porpoise's rostrum (when the beam was broken), and 'absence'

when the porpoise's rostrum was outside the optical beam. Significant effort was put into fine-tuning the dimensions so the interval between the start of the response and the moment that the sensor indicated 'absence' (i.e. the motor component of the response) was minimal, without false detections. The sensor was cleaned daily to prevent algal growth that would have influenced the measurements.

The reaction time sensor communicated via binary electrical signals with the DAQ device, which was controlled by a custom-written MATLAB program. The program allowed the operator during research sessions to set the stimulus level and measure the animal's RT [defined here as the interval between the trigger of the test signal (which was loaded into the computer memory before triggering) and the moment the animal moved out of the optical beam]. The output of the sensors was sampled real-time at a rate of 125 Hz (8 ms resolution). This rate was the maximum rate possible to achieve stable sampling, which was verified before each research session by simulation of a test trial.

A second 319 THz infrared LED in the top sensor bracket allowed the signal operator to check whether the reaction time sensor was working correctly. The LED was switched on automatically when the animal was present, and was captured by an underwater camera (Mariscope Micro, Puerto Montt, Chile) filming the listening station from above (Fig. 4.1). The underwater camera made the infrared light visible on the monitor image. The images from the camera, together with the sound from a microphone inside the research cabin, were digitized by using a video analog-to-digital converter (Geniatech EZ Grabber, Shenzhen, China) and shown on a laptop screen to the signal operator during research sessions. The images were also visible to the trainer on a monitor near the start/response buoy.

Calibration of sound stimuli

The sound calibration equipment consisted of two hydrophones (B&K 8106) with a multichannel high frequency analyser (B&K PULSE 3560 D), and a laptop computer with B&K PULSE software (Labshop version 12.1). The system was calibrated with a pistonphone (B&K 4223). The received SPL of each test signal was derived from the

90% energy flux density, divided by the corresponding 90% time duration (Madsen, 2005).

The background noise levels were measured multiple times, under research session conditions: water and air circulation system off, no rain, and wind force Beaufort 4 or below. 1/3-Octave band SPLs of the background noise were determined by averaging the squared sound pressure in the 100 Hz to 160 kHz bands over a period of 10 s. During calibration measurements the background noise in the pool was very low; above 3.5 kHz it was just above the self-noise of the recording equipment.

The received SPL of each test signal was measured once or twice (depending on the frequency). These measurements were conducted using the two hydrophones, one at each location of the auditory meatus of the porpoise when he was positioned at the listening station. The SPL at the two locations differed by 0-7 dB (mean absolute difference 3 dB). After averaging of the SPL over the two hydrophone locations, the difference in average SPL between measurement days was 1-3 dB (depending on the frequency). The final calibration value was taken as the grand mean over the hydrophone locations and measurement days.

Received SPLs were calibrated using relative output levels of 60-100 dB. The linearity of the transmitter system was checked at 0.5, 1, and 4 kHz; it was consistent to 1 dB within the 40 dB range.

Experimental procedure

A trial began when the porpoise touched the start/response buoy with his rostrum. When the trainer gave a vocal command and pointed downwards, the porpoise swam to the listening station (Fig. 4.4A) and positioned his rostrum against it, so that his anterior-posterior axis was aligned with the acoustic beam axis of the projector (Fig. 4.4B). Using the images from the underwater camera, the trainer judged whether or not the animal was positioned correctly. If he was, the trial would continue; if he was not, the trainer knocked on the start/response buoy, the porpoise returned to the buoy, and the trainer sent him straight back to the listening station. Once positioned at the listening

station, the porpoise was trained to respond (Fig. 4.4C) upon detecting either the test stimulus or the trainer's whistle by returning to the start/response buoy (Fig. 4.4D), and to stay at the listening station until he heard a signal.

Research sessions consisted of 75% signal-present trials and 25% signal-absent (or 'catch') trials. In all trials the porpoise waited at the listening station for a random period between 4 and 10 s. In signal-present trials, the signal operator played the test signal from the custom-written MATLAB program after the random waiting time. When the sound was being transmitted, a video distorter produced horizontal lines in the video image (Fig. 4.4C), which helped the operator to determine whether or not the porpoise had responded to the test sound. If the animal responded within 2 s of signal onset, the operator indicated to the trainer that the response was correct using a hand gesture, after which the trainer gave the porpoise a fish reward. If the animal did not respond within 2 s, the operator signalled to the trainer that the trial had ended. The trainer then called the porpoise back to the start/response buoy by softly tapping three times on the side of the pool, and no fish reward was given. In signal-absent trials, the porpoise stationed, and after the random waiting time the operator gestured to the trainer to either blow on a whistle or to softly tap three times on the side of the pool (in relative proportions of 1:1). For returning to the start/response buoy directly after a whistle, the animal also received a fish reward. The trainer did not know beforehand whether a trial was a signal-present or signal-absent trial.

If the animal responded before a signal was produced (pre-stimulus response), the signal operator indicated this to the trainer who then ignored the animal for about 10 s before starting a new trial. Pre-stimulus responses were ignored when they were clearly initiated by external sounds; sessions continued as soon as the sound had stopped.

An experimental session consisted of 30-35 trials and lasted for about 20 min. For each session, one of four data collection sheets was used; each sheet had a random series of waiting times and a balanced number of trials per signal level. The signal levels were randomized, with the restriction that the level difference between successive trials was not more than 30 dB (*sensu* Wagner et al., 2004).

Research sessions were conducted in May to July 2011 and in August and September 2012. Three experimental sessions per day were conducted five days a week in 2011 (sessions started at 0900, 1100, and 1400 h), and one extra session was performed daily in 2012 (starting at 1600 h). In 2011, test frequencies ranged from 4 to 125 kHz, and on average 39 RT measurements were collected per level/frequency combination. The test frequency was changed from day to day and adjacent frequencies were usually tested on successive days (going from high to low and from low to high frequencies). In 2012, RTs for frequencies of 0.5, 1, and 2 kHz were also measured, and the existing datasets for other frequencies were increased until at least 50 RT measurements per level/frequency combination were available to calculate the equal latency contours.

Data processing and analysis

Medians were calculated rather than means, because the distributions of RTs, especially for levels near the hearing threshold, were often skewed. The relationship between mean or median RT and stimulus intensity has been described by a decaying power law approaching an asymptote at high intensities (Luce, 1986; Piéron, 1920). Initial fits of this three-parameter model, known as Piéron's law, to the median RTs of the porpoise provided reasonable approximations to the data. However, the estimate of the asymptotic parameter, which is suggested to reflect a minimum processing time and motor component (Luce and Green, 1972), was unstable and often became negative. Therefore, a two-parameter power law was fitted to the median RTs (in ms) for each test frequency:

$$\mathrm{RT} = \beta(I/I_0)^{-\alpha}\,, \qquad (4.1)$$

where I/I_0 is the ratio of the intensity of the test stimulus (I) to the intensity of a stimulus at threshold (1_0) (calculated using $I/I_0 = 10^{\mathrm{SnL}/10}$), exponent α is the slope on a log-log scale, and β is the y-intercept (equal to the RT at SnL = 0 dB). This two-parameter function was less sensitive to variation in the median RT and gave similar results in terms of goodness of fit as Piéron's law. All fits of the auditory RT functions were made using a non-linear method (Trust Region algorithm) in MATLAB.

For levels near the hearing threshold, statistical measures of RT are affected by the animal's response criterion, i.e. the animal's tendency to give a positive or negative response (Heil et al., 2006), and relatively long RTs often occur that result in deviation from simple power law behaviour (Pins and Bonnet, 2000; Stebbins and Miller, 1964; Wagner et al., 2004). Therefore, one or two median RTs (depending on the frequency) to low intensity signals (SnL < 30 dB) were omitted when this substantially improved the model fits (omitted data are shown in Fig. 4.5). Finally, the best-fitting auditory RT models were evaluated at reference RTs of 150, 160, 170, 180, 190, and 200 ms to determine the SnLs (and, hence, the SPLs) of the equal latency contours (labelled I-VI, respectively). These reference values were selected because, except for one data point at 16 kHz, the SPLs of the six contours always fell within the range of tested levels.

To derive six auditory weighting functions from the equal latency contours, the data sets were adapted and smoothed using the shape of the animal's own audiogram as a template. The rationale behind this approach was as follows: 1) smoothing was justified because the range of RTs was small, and weighting functions are generally idealized curves, 2) the audiogram of the subject had been determined very accurately, and was similar to that of two other harbour porpoises over most of the hearing range (Andersen, 1970; Kastelein et al., 2002; Kastelein et al., 2010); 3) the equal loudness contours and audiogram were expected to have similar shapes but to have different low frequency roll-off rates; and 4) the equal latency and equal loudness contours were expected to have similar shapes, except possibly at very high frequencies.

Each smoothed contour (L_{lat}) was a transformation of the hearing threshold (L_{ht}):

$$L_{\text{lat}}(f) = \gamma L_{\text{ht}}(f) + \delta, \tag{4.2}$$

where f is a vector of test frequencies, and γ and δ are a scaling and translation parameter, respectively. The root-mean-square error (RMSE) between the SPLs of the equal latency contour and the SPLs of the transformed hearing threshold was calculated for combinations of γ and δ using a simple iterative algorithm. The RMSE indicated the

similarity between the two curves; the minimum RMSE determined the combination of γ and δ that corresponded to the best-fitting function (the 'smoothed' contour). This process was repeated five times for each of the six contours to investigate the influence of the high frequency data on the similarity; once with the full data set (all frequencies) and four times with part of the data set (<31.5 kHz, <63 kHz, <80 kHz, and <125 kHz).

Exclusions of high frequency data did not have clear effects on the similarity between the unsmoothed and smoothed versions of contours V and VI; the smallest RMSE was ~5 dB for contour V and ~6 dB for contour VI, independent of the range of frequencies included (Table 4.1). However, the similarity between the unsmoothed and smoothed versions of contours I-IV increased significantly after the exclusion of high frequency data (Table 4.1), and the best results (smallest RMSEs) were obtained when 63, 80, and 125 kHz were omitted. The decreased similarity was suspected to be due to a weak RT-loudness correlation (see Discussion), so only the smoothed 0.5-31.5 kHz data sets were used in further analyses. For these data sets, the best-fit estimates were: for parameter γ were: 0.610, 0.721, 0.825, 0.924, 1.016 and 1.104 dB/dB, and for parameter δ were: 103.77, 85.94, 69.22, 53.39, 38.52 and 24.37 dB, for contours I-VI, respectively.

The smoothed contour data sets were extended with frequencies of 0.25, 8, and 50-150 kHz by using the porpoise's own hearing thresholds for these frequencies (Table 4.2) (Kastelein et al., 2010) in the calculated contour/threshold relationships (Eqn 4.2). This resulted in a family of six hypothetical equal loudness contours. A closed-form model was fitted to each of these contours:

$$L_{\text{loud}}(f) = K_1 - 20\log_{10}\left[\frac{b^x f^x}{\left(a^x + f^x\right)\left(b^x + f^x\right)}\right], \qquad (4.3)$$

where L_{loud} is the SPL of the equal loudness contour in dB re 1 μPa, f is the frequency in Hz, and K_1, a, b, and x are parameters that determine the shape of the function. The mathematical form of the C- and M-weightings (Kinsler, 2000; Southall et al., 2007) also used by Finneran and Schlundt (2011) is a special case of Eqn 4.3 where x is 2. Here, parameter x was fitted as a free parameter to allow the roll-off rate to vary.

Parameters *a* and *b* in Eqn 4.3 represent the lower and upper frequency where the level is 6 dB above the minimum of the curve, respectively. Parameter *b* was set to 10^9 Hz so that the high frequency roll-off was effectively non-existent within the frequency range of hearing.

The best-fitting functions of Eqn 4.3 were normalized to 0 dB at the most sensitive frequency and inverted to obtain the auditory weighting functions, which then take the form:

$$W(f) = K_2 + 20\log_{10}\left[\frac{b^x f^x}{\left(a^x + f^x\right)\left(b^x + f^x\right)}\right], \qquad (4.4)$$

where $W(f)$ is the weighting level in dB, a, b, and x are as in Eqn 4.3, and K_2 is a vertical offset that results from the normalization of the contours ($K_2 = \min\{L_{loud}\} - K_1$).

Comparison of contours with TTS onset and behavioural response onset thresholds

To get an indication of which of the weighting functions is the most relevant for predicting the onset of behavioural responses and temporary hearing loss, the hypothetical equal loudness contours were compared to two data sets derived from the literature. The results of this exploratory analysis are preliminary as it is based on a very small data set. The data sets used are listed below:

1. TTS onset thresholds of a harbour porpoise after exposure to 1-2 kHz or 6-7 kHz sonar signals (Kastelein et al., 2014, 2015b) or octave-band noise centred at 4 kHz (Kastelein et al., 2012a). A TTS onset threshold was defined as the SEL_{cum} of the exposure sound that is needed to induce a 6 dB threshold shift when hearing is measured within 4 minutes after exposure (Finneran, 2015). The onset thresholds were averages based on different combinations of SPL, duty cycle and exposure duration. Two subsets were created from the sonar data, for 10% and 100% duty cycles respectively, as this parameter was shown to have a strong effect on the magnitude of the induced TTS (Kastelein et al.,

2014). The study subject (ID 02) in the three TTS studies was the same as in the present study. Data from other studies on TTS in porpoises were excluded from the analysis because they either only induced higher levels of TTS (>20 dB) immediately after exposure (Popov et al., 2011) or used a very different type of exposure signal; that is, a single seismic airgun shot with a relatively broad frequency spectrum (Lucke et al., 2009).

2. Behavioural response thresholds of free-ranging and captive harbour porpoises in response to various types of signals. Tougaard et al. (2015) derived onset thresholds for avoidance behaviour from 10 published field studies in which harbour porpoises were exposed to pingers, seal scarers or pile driving sounds. To make the data comparable across studies, the authors specified the onset thresholds in terms of $L_{eq\text{-}fast}$, which is equivalent to a "fast-average" SPL with a RMS averaging time of 125 ms. These thresholds were taken from table 1 in Tougaard et al. (2015) and combined with behavioural response thresholds for harbour porpoises derived from captive studies. Specifically, Kastelein et al. (2005b; 2008a; 2008b) used the average animal swimming location and the sound gradient in the floating pen to determine avoidance (or, "discomfort") thresholds in response to pure tones and underwater communication signals; Kastelein et al. (2012b; 2013b) measured the proportion of occurrence of brief changes in speed and/or direction (described in the earlier study as "startle responses") in response to sonar signals in the 1-7 kHz frequency band; and Kastelein et al. (2015a) measured increases in respiration rate and increases in swimming speed in response to 25 kHz sonar signals. To enable comparison with the field data, these response thresholds from captive studies were specified in terms of $L_{eq\text{-}fast}$ (see Table 4.3 for more details on the captive data).

The six contours were independently fitted to both the TTS onset and behavioural response data sets described above. There are indications that lower amounts of SEL_{cum} are needed to induce TTS using exposures to band noise signals compared to exposures to tonal signals, which are usually of relatively low SPL and long duration (Finneran, 2015). The contours were therefore also fitted to a data set that did not include the

thresholds for 4 kHz octave-band noise. For frequencies >20 kHz, the behavioural response data showed unusually large variation over a range of ~60 dB and divergence was observed between the thresholds from captive and free-range studies. One possible explanation for this increased variation is that the hearing of porpoises becomes increasingly more directional towards these higher frequencies (Kastelein et al., 2005a). As the low frequency roll-off rates of the curves were eventually of most interest, data for frequencies >20 kHz were excluded from the fit.

In the fitting process, the shape of the contour was maintained and only the vertical offset of the contour was allowed to vary. Thus, only parameter K_1 in Eqn 4.3 was re-estimated and the shape parameters *a*, *b* and *x* were fixed at the values derived from fitting the equal-latency data. Goodness-of-fit measures r^2 and RMSE were used to quantify the correspondence between the vertically offset contours and the TTS onset or behaviour response onset data.

To visualise the effect of wind-driven ambient noise on the detection of signals by harbour porpoises, the frequency-specific noise-limited theoretical detection threshold (DT; in dB re 1 μPa; Ellison et al., 2012) was calculated for 6 sea states:

$$DT(f) = NL(f) + CR(f) - DI(f), \quad (4.5)$$

where NL is the power spectrum density level (in dB re 1 $\mu Pa^2\ Hz^{-1}$) of the noise given a sea state, based on Knudsen (1948), CR is the critical ratio (in dB re 1 Hz) for the harbour porpoise (Kastelein et al., 2009), and DI is the receiving directivity index (in dB; Kastelein et al., 2005). These curves represent the SPL at which there is a 50% probability that a pure tone is heard during average wind-driven ambient noise conditions, given the assumptions of Gaussian background noise and a random swimming direction of the porpoise. More realistic predictions can be made for non-Gaussian noise that is modulated in amplitude across frequencies by including a correction based on the co-modulation index (Branstetter et al., 2013).

RESULTS

A total of 5,144 trials were conducted in 167 experimental sessions, resulting in 3,822 RT measurements. Only 28 pre-stimulus responses occurred throughout the study (0.5% of the total number of trials), of which 17 occurred during the first five sessions (when the animal was still getting used to the test procedure).

The median observed RTs of the harbour porpoise to the nine FM tonal signals are shown in Fig. 4.5 as functions of both SnL and SPL. The porpoise responded after signal onset with median RTs of between 95 and 522 ms. RT decreased with increasing SPL at every frequency. The auditory RT functions fitted to the median RTs generally exhibited the steepest log-log slopes (α closer to unity) at the lower and higher frequencies, and increasingly showed shallow slopes towards the middle frequencies (Table 4.2; Fig. 4.5). The goodness of fit values were satisfactory: the coefficient of determination (r^2) ranged from 0.90 to 0.99 and the RMSE ranged from 2.7 to 6.4 ms (Table 4.2).

Six equal latency contours (I-VI) were constructed from the auditory RT functions. The two lowest equal latency contours (V and VI; corresponding to 190 and 200 ms, respectively) roughly followed the shape of the hearing threshold, including the notch in the audiogram at 63 kHz (Fig. 4.6). On average, the audiogram and contour VI were 31 dB apart (range 20-41 dB). The average spacing between adjacent equal latency contours was greater in the mid-range (16-31.5 kHz; 11-13 dB) than in the low range (0.5-4 kHz; 6-9 dB) and high range of test frequencies (63-125 kHz; 5-8 dB), an effect that directly relates to the slopes of the auditory RT functions (Table 4.2**;** parameter α).

The six equal latency contours were converted into hypothetical equal loudness contours (Fig. 4.7) and auditory weighting functions (Fig. 4.8); the parameter estimates for Eqns 4.3 and 4.4 are provided in Table 4.4. The weighting level at the lowest frequency (250 Hz) was between –72 dB and –41 dB, depending on the weighting function. The –6 dB point and –3 dB point matched a frequency between 4.6 and 5.9 kHz and between 7.8 and 8.2 kHz, respectively; the weighting level was 0 dB for frequencies of ≥17.1 to ≥25.2 kHz. The low frequency roll-off rate of the weighting

function ranged from 32 to 51 dB/decade (9.7 to 15.9 dB/octave), depending on the equal latency contour it was based upon.

Both the TTS onset (Fig. 4.9A) and behavioural response onset (Fig. 4.9B) thresholds generally decreased as functions of peak signal frequency for frequencies below 20 kHz. The TTS data for tonal and band noise signals combined were best described by the shape of contour II (see Table 4.5 for r^2 and RMSE values). However, the shape of contour I most accurately approximated the reduced data set of only TTS onset data for 1-2 and 6-7 kHz sonar signals. Despite the very low sample size, these preliminary results tentatively suggest that the TTS onset thresholds are best described by a weighting function with a relatively shallow low frequency slope.

Compared to the TTS data, the behavioural response onset thresholds derived from field and captive studies were most accurately described by a contour with a greater low frequency roll-off. Contours IV and V had the highest r^2 values out of all six contours (0.86 for both; Table 4.5). Contour V and the animal's own audiogram had an almost identical low frequency roll-off (parameter γ was estimated as 1.0 for this contour); therefore, based on these data, it seems appropriate to use the audiogram as a weighting function for behavioural effects in the absence of alternative methods. The average difference between the audiogram and the fit of contour V to the behavioural thresholds was approximately 50 dB.

DISCUSSION

Methodological considerations

The hearing abilities of the study animal were probably representative for porpoises of his age and younger, as his hearing thresholds under unmasked and masked conditions measured 1.5 to 5 years earlier were similar to those of two other male harbour porpoises (Kastelein et al., 2002; Kastelein et al., 2009; Kastelein et al., 2010). The auditory weighting functions (Fig. 4.8) were based on the equal latency contours and the hearing thresholds of the animal; therefore, these functions may also be representative for other members of the species.

The six reference RTs of 150 to 200 ms were chosen to simplify the interpretation of the results. When equal latency contours are averaged across subjects, it is more accurate to use one reference frequency at which, for each individual, the reference RTs of the contours are determined that match predefined sensation levels (Pfingst et al., 1975a). This approach reduces the between-subject variation in RT that commonly occurs (e.g., Epstein and Florentine, 2006; Humes and Ahlstrom, 1984), particularly if this variation is frequency independent.

Very few pre-stimulus responses occurred, which shows that the porpoise mainly refrained from guessing, probably because most of the levels were well above the animal's hearing threshold. The animal was not trained to respond as quickly as possible, so the RTs found here might represent conservative estimates. However, the porpoise's RTs were probably shorter than its species' average because the animal was highly experienced in stimulus detection tasks (Blackwood, 2003). The higher pre-stimulus response rate during the first five sessions was probably because the animal had to get used to the new procedure.

The difference in SPL between the two outer equal latency contours (I and VI) was as high as 67 dB. Similar differences are common in humans and other species for medium and high SnLs (Luce, 1986; Stebbins, 1966). The data collection protocol was designed to provide a large enough sample to capture the decline in RT with increasing SPL, at sufficient frequencies (nine) to cover the wide hearing range of the animal. The lowest test frequency was 500 Hz. Acoustic calibrations of 250 and 400 Hz sound signals showed that harmonics occurred at levels judged to be too close to the hearing threshold at these frequencies, despite the fact that the low frequency projector was one of the most powerful non-military sources available.

This study was focused upon mid-range and high-range SnLs; therefore, the hearing thresholds of the harbour porpoise were not re-evaluated during the study and relatively few test signals had low SnLs—insufficient to inform the auditory RT model with four parameters presented by Wagner et al. (2004). For SnLs between 0 and 40 dB, plotting

the median RTs for all frequencies in a similar graph as Fig. 4.5 showed that, despite the differences in minimum level, the 0.5 to 80 kHz functions were very similar. This suggests that the equal latency contours of the porpoise closely follow the shape of the hearing threshold at low SnLs, as expected from the equal loudness and equal latency contours of humans (Chochelle, 1940; Suzuki and Takeshima, 2004).

For tests in the 125 kHz band, the median RTs near the hearing threshold of the porpoise differed significantly from those for other frequencies. At the lowest test level (SnL = 18 dB) the median was 522 ms. This value was expected to be much closer to the hearing threshold level, especially for a small odontocete like the harbour porpoise (Blackwood, 2003). Click rates were sometimes heard by the signal operator through the monitoring system before and during presentation of the 125 kHz signals, and it is therefore possible that some test signals were not audible to the porpoise because his echolocation click trains masked detection of the signals. This may also explain the slight increase in the equal latency contour values relative to 80 kHz. Click trains were not heard when frequencies below 125 kHz were tested. A re-evaluation of the subject's hearing thresholds was recently performed which showed no substantial changes in the audiogram over a 3–4 year period (Kastelein et al., 2013a). In this more recent study, the porpoise was not allowed to echolocate during research trials.

Relationship between reaction time and loudness

In humans, simple RTs correlate with direct estimates of loudness (Luce, 1986; Marks and Florentine, 2011), and RT is often used as a proxy measure of loudness (Arieh and Marks, 2003; Florentine et al., 2004; Wagner et al., 2004). RT has therefore been used in subjects for which loudness assessment with standard methods is very difficult or impossible, such as human infants (Leibold and Werner, 2002) and non-human animals (Dooling et al., 1978; Green, 1975; Kastelein et al., 2011b; May et al., 2009; Moody, 1973; Pfingst et al., 1975a; Stebbins, 1966; Ridgway et al., 2001). Functionally, the RT reflects the combined duration of the sensory, cognitive and motor processes needed to generate the response (Saunders, 1998). RT is not determined by properties of the received sound stimulus alone but also by, for example, age (Birren and Botwinick,

1955), body size (Blackwood, 2003), and masking noise levels (Chocholle and Greenbaum, 1966).

The relationship between loudness (derived from magnitude estimation) and sensation level above ~30 dB is best described by a simple power law that is almost identical to Eqn 4.1 (Stevens, 1955). At these moderate to high levels, the slopes of such loudness functions are negatively correlated with the slopes of auditory RT functions for the individual listener (Humes and Ahlstrom, 1984; Reason, 1972). For SnLs lower than ~30 dB, both the loudness function and the auditory RT function diverge from this simple power law (Chocholle, 1940; Hellman and Zwislocki, 1961; Takashima, 2003).

Most researchers investigating the relationship between RT and loudness have used only one or two test frequencies in the range of most sensitive hearing. When more test frequencies are used, slopes of loudness and auditory RT functions are frequency dependent at moderate to high SnLs, and equal latency and equal loudness contours are similar in shape (Chocholle, 1940; Marshall and Brandt, 1980; Pfingst et al., 1975a). In general, RTs do not vary with frequency at low SnLs, so that equal latency contours follow the shape of the hearing threshold, although deviations have been reported for some listeners (Epstein and Florentine, 2006). Kohfeld et al. (1981) also reported a discrepancy between equal latency contours and equal loudness contours at lower levels (20 and 40 phons), but this was later attributed to the loudness-matching procedure that was used (Buus et al., 1982).

There is a negative relationship between the increase in perceived loudness with SnL and the spacing between equal loudness contours (i.e. loudness increases more steeply with SnL where the spacing between the contours is smaller and *vice versa*). In humans, smaller increases in loudness with SnL are observed at frequencies within the range of most sensitive hearing than at lower frequencies, which causes the contours to flatten towards higher loudness levels (Suzuki and Takeshima, 2004). In this study, the equal latency contours of the porpoise showed a similar trend for frequencies up to 31.5 kHz, which suggests a strong correlation between RT and loudness at these frequencies.

Less spacing between the equal latency contours was observed not only for low frequencies but also for frequencies of 63, 80, and 125 kHz (Fig. 4.6), an effect that cannot be expected based on the equal loudness contours of humans (Suzuki and Takeshima, 2004) or of a bottlenose dolphin (Finneran and Schlundt, 2011). If tones of equal loudness truly elicit equal RTs in harbour porpoises, then the results indicate that the dynamic hearing range of the porpoise is very narrow at these high frequencies. Harbour porpoise echolocation clicks contain sound energy mainly at frequencies of 110-150 kHz (Møhl and Andersen, 1973); therefore, a narrow dynamic hearing range at these frequencies seems unlikely. The animals encounter large differences in SPL at these frequencies in their daily life; they experience very faint echoes of their own echolocation clicks and high intensity clicks (peak-to-peak source levels: 178-205 dB re 1 μPa m; Villadsgaard et al., 2007) from other porpoises. The harbour porpoise inner ear has an acoustic fovea on the basilar membrane with high ganglion cell densities in the region where these echolocation frequencies are processed (Ketten, 1997); the relatively short RTs that were found in this study may have been the result of increased neural activity generated in the foveal region. In addition, the RTs of other mammalian species for frequencies higher than ~16 kHz reported by Green (1975), May et al. (2009) and Kastelein et al. (2011b) were also shorter than expected from the equal loudness contours measured to date, suggesting that the correlation between RT and loudness is consistently weaker at these very high frequencies.

Conclusions and recommendations

The six auditory weighting functions (Fig. 4.8) are assumed to represent relative loudness perception in the porpoise. The experimental method used in this study is relatively fast compared to direct loudness estimation, and could be applied to any species that can be trained to perform psychophysical go/no-go tasks. However, there is currently only indirect evidence in favour of the weighting method based on equal latency. A direct comparison between equal latency and equal loudness contours over a wide range of frequencies in the same subject would minimize the uncertainty in the outcome that results from the assumptions of 1) a strong relationship between loudness and RT at low and middle frequencies, and 2) divergence from this relationship at very high frequencies.

The flattest weighting function (curve I in Fig. 4.8) is associated with the loudest sounds. TTS is generally induced by loud sounds; therefore, a function relating TTS onset thresholds to frequency is expected to be similar in shape to the flattest equal loudness contour. The very few relevant TTS onset data available for the harbour porpoise (Kastelein et al., 2012a, 2014, 2015b) agree with this expectation (Fig 9A). However, a weighting function based on a lower equal loudness contour, thus with relatively more curvature, predicted TTS onset levels most accurately in bottlenose dolphins (Finneran and Schlundt, 2013). Similarly, exposures to half-octave band noise at centre frequencies of 22, 32, 45 and 90 kHz induced relatively high levels of TTS (~15-30 dB) at received SELs of 161-165 dB re 1 μPa s in two Yangtze finless porpoises, *Neophocaena phocaenoides asiaeorientialis* (Popov et al., 2011; Finneran, 2015); thus, these levels are significantly lower than predicted by the curve fits for TTS onset in Fig. 4.9A. This may suggest that TTS onset levels are not always perceived as equally loud across frequencies, and indicates a discrepancy that should be re-evaluated once more TTS onset thresholds become available.

For frequencies below 20 kHz, there seems to be a good correlation between the onset of a behavioural response in harbour porpoises and the sensation level (SnL) of the exposure signal, with responses generally occurring at levels of around 50 dB above the hearing threshold (Fig 9B). Part of the data set used here came from Tougaard et al. (2015), so it is perhaps unsurprising that these authors reached similar conclusions. The patterns in response thresholds were comparable between data collected in laboratory and free-ranging conditions, at least for frequencies <20 kHz. Despite this apparent similarity in behaviour response onset in the harbour porpoise, it is important to note that behavioural responses observed in captivity should not be assumed *a priori* to be representative for wild animals, as the motivations and underlying cognitive processes leading to the change in behaviour might be very different. In addition, the captive thresholds used here came from 4 animals which introduced a degree of pseudoreplication in the data set (Table 4.3; Kastelein et al., 2005b, 2012b, 2013b, 2015a). In many of these laboratory studies, the research objective was to determine the relative effect of the signal characteristics on behaviour, and a substantial part of the

variation in behavioural thresholds was ascribed to these different signal characteristics. However, variation in the behavioural thresholds obtained from the field studies may have also been caused differences in signal types, but were likely also influenced by other factors such as ambient noise conditions, age, motivational state, and prior experience with the sound stimulus (Ellison et al., 2012).

Acoustic safety criteria for the exposure of marine mammals to anthropogenic noise can be made more accurate with auditory weighing functions such as those obtained in the present study, because behavioural and physiological responses of marine mammals to noise are expected to correlate better with the perceived loudness of a sound than with the unweighted SPL (Finneran and Schlundt, 2013; Southall et al., 2007). Frequency weighting based on equal latency may also help to determine whether the current noise safety regulations are appropriate. These regulations may, for instance, be too conservative for low frequency signals, and too liberal for high-frequency signals, or *vice versa*. At the intermediate SPLs which are needed to trigger behavioural responses in harbour porpoises, the difference in shape between equal loudness and equal-SnL contours is very small and both are expected to predict behavioural response thresholds equally well (or poorly). However, for physiological responses the difference between equal loudness and sensation level might be more pronounced. Indeed, the preliminary analysis of TTS data seems to suggest that loudness/latency can be a more relevant predictor of TTS onset; however this preliminary conclusion is based on a very limited data set. The lack of information about the influence of signal frequency on TTS onset thresholds, not only for harbour porpoises but for marine mammals in general, clearly represents a critical data gap (Finneran, 2015).

The weighting functions in this study may be used to predict behavioural response thresholds independent of the frequency of the signal that caused the response, from measured behavioural response thresholds to signals of known frequency spectra. Such extrapolations would greatly increase the applicability of behavioural response thresholds for marine mammals in the wild that were measured during exposure to relatively narrowband sources (e.g., Miller et al., 2012; Tyack et al., 2011). The weighting functions may also enable more accurate estimations of the distances from a

variety of sound sources at which behavioural responses, such as avoidance of the sound source, and physiological responses, such as the onset of TTS, occur in marine mammals.

REFERENCES

Andersen, S. (1970). Auditory sensitivity of the harbour porpoise *Phocoena phocoena*. In Investigations on Cetacea (ed. G. Pilleri), pp. 255-259. Berne, Switzerland: Berne Brain Anatomy Institute.

Arieh, Y. and Marks, L. E. (2003). Recalibrating the auditory system: a speed-accuracy analysis of intensity perception. J. Exp. Psychol. Hum. Percept. Perform. 29, 523-536.

Birren, J. E. and Botwinick, J. (1955). Speed of response as a function of perceptual difficulty and age. J. Gerontol. 10, 433-436.

Blackwood, D. J. (2003). Vocal response times to acoustic stimuli in white whales and bottlenose dolphins. PhD thesis, Texas A&M University, College Station, USA.

Branstetter, B. K., Trickey, J. S., Aihara, H., Finneran, J. J. and Liberman, T. R. (2013). Time and frequency metrics related to auditory masking of a 10 kHz tone in bottlenose dolphins (*Tursiops truncatus*). J. Acoust. Soc. Am. 134, 4556-4565.

Buus, S., Greenbaum, H. and Scharf, B. (1982). Measurements of equal loudness and reaction times. J. Acoust. Soc. Am. 72, S94.

Chocholle, R. (1940). Variation des temps de réaction auditifs en fonction de l'intensité à diverses fréquences. Ann. Psychol. 41, 65-124.

Chocholle, R. and Greenbaum, H. (1966). La sonie de sons purs partiellement masqués Étude comparatif par une méthode d'égalisation et par la méthode des temps de réaction. J. Physiol. 63, 386-414.

Chowning, J. M. (1973). The synthesis of complex audio spectra by means of frequency modulation. J. Audio Eng. Soc. 21, 526-534.

De Jong, C. A. F. and Ainslie, M. A. (2008). Underwater radiated noise due to the piling for the Q7 Offshore Wind Park. Proceedings of Acoustics '08, Paris, France.

Dooling, R. J., Zoloth, S. R. and Baylis, J. R. (1978). Auditory sensitivity, equal loudness, temporal resolving power, and vocalizations in the house finch (*Carpodacus mexicanus*). J. Comp. Physiol. Psychol. 92, 867-876.

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.

Epstein, M. and Florentine, M. (2006). Reaction time to 1- and 4-kHz tones as a function of sensation level in listeners with normal hearing. Ear Hear. 27, 424-429.

Finneran, J. J. (2015). Noise-induced hearing loss in marine mammals: A review of temporary threshold shift studies from 1996 to 2015. J. Acoust. Soc. Am. 138, 1702-1726.

Finneran, J. J. and Schlundt, C. E. (2007). Underwater sound pressure variation and bottlenose dolphin (*Tursiops truncatus*) hearing thresholds in a small pool. J. Acoust. Soc. Am. 122, 606-614.

Finneran, J. J. and Schlundt, C. E. (2011). Subjective loudness level measurements and equal loudness contours in a bottlenose dolphin (*Tursiops truncatus*). J. Acoust. Soc. Am. 130, 3124-3136.

Finneran, J. J. and Schlundt, C. E. (2013). Effects of fatiguing tone frequency on temporary threshold shift in bottlenose dolphin (*Tursiops truncatus*). J. Acoust. Soc. Am. 133, 1819-1826.

Fleshler, M. (1965). Adequate acoustic stimulus for startle reaction in the rat. J. Comp. Physiol. Psychol. 60, 200-207.

Fletcher, H. and Munson, W. A. (1933). Loudness, its definition, measurement and calibration. J. Acoust. Soc. Am. 5, 82-108.

Florentine, M., Buus, S. and Rosenberg, M. (2004). Reaction-time data support the existence of Softness Imperfection in cochlear hearing loss. In Auditory Signal Processing: Physiology, Psychoacoustics, and Models (ed. D. Pressnitzer, A. de Cheveigné, S. McAdams and L. Collet), pp. 28-34. New York, NY: Springer Verlag.

Götz, T. and Janik, V. M. (2011). Repeated elicitation of the acoustic startle reflex leads to sensitisation in subsequent avoidance behaviour and induces fear conditioning. BMC Neurosci. 12, 30.

Green, S. (1975). Auditory sensitivity and equal loudness in the squirrel monkey (*Saimiri sciureus*). J. Exp. Anal. Behav. 23, 255-264.

Heil, P., Neubauer, H., Tiefenau, A. and von Specht, H. (2006). Comparison of absolute thresholds derived from an adaptive forced-choice procedure and from reaction

probabilities and reaction times in a simple reaction time paradigm. J. Assoc. Res. Otolaryngol. 7, 279-298.

Hellman, R. P. and Zwislocki, J. (1961). Some factors affecting the estimation of loudness. J. Acoust. Soc. Am. 33, 687-694.

Humes, L. E. and Ahlstrom, J. B. (1984). Relation between reaction time and loudness. J. Speech Hear. Res. 27, 306-310.

Kastelein, R. A., van den Belt, I., Helder-Hoek, L., Gransier, R. and Johansson, T. (2015a). Behavioral Responses of a Harbor Porpoise (*Phocoena phocoena*) to 25-kHz FM Sonar Signals. Aquat. Mamm. 41, 311-326.

Kastelein, R. A., Bunskoek, P., Hagedoorn, M., Au, W. W. L. and de Haan, D. (2002). Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals. J. Acoust. Soc. Am. 112, 334–344.

Kastelein, R. A., Gransier, R., Hoek, L. and Olthuis, J. (2012a). Temporary threshold shifts and recovery in a harbor porpoise (*Phocoena phocoena*) after octave-band noise at 4 kHz. J. Acoust. Soc. Am. 132, 3525-3537.

Kastelein, R. A., Gransier, R., Hoek, L. and Rambags, M. (2013a). Hearing frequency thresholds of a harbor porpoise (*Phocoena phocoena*) temporarily affected by a continuous 1.5 kHz tone. J. Acoust. Soc. Am. 134, 2286-2292.

Kastelein, R. A., Gransier, R., van den Hoogen, M. and Hoek, L. (2013b). Brief behavioral response threshold levels of a harbor porpoise (*Phocoena phocoena*) to five helicopter dipping sonar signals (133 to 143 kHz). Aquat. Mamm. 39, 162-173.

Kastelein, R. A., Gransier, R., Schop, J. and Hoek, L. (2015b). Effects of exposure to intermittent and continuous 6–7 kHz sonar sweeps on harbor porpoise (*Phocoena phocoena*) hearing. J. Acoust. Soc. Am. 137, 1623-1633.

Kastelein, R. A., Hoek, L., Gransier, R., Rambags, M. and Claeys, N. (2014). Effect of level, duration, and inter-pulse interval of 1-2 kHz sonar signal exposures on harbor porpoise hearing. J. Acoust. Soc. Am. 136, 412-422.

Kastelein, R. A., Hoek, L. and de Jong, C. A. F. (2011a). Hearing thresholds of a harbor porpoise (*Phocoena phocoena*) for sweeps (1-2 kHz and 6-7 kHz bands) mimicking naval sonar signals. J. Acoust. Soc. Am. 129, 3393-3399.

Kastelein, R. A., Hoek, L., de Jong, C. A. F. and Wensveen, P. J. (2010). The effect of signal duration on the underwater detection thresholds of a harbor porpoise (*Phocoena phocoena*) for single frequency-modulated tonal signals between 025 and 160 kHz. J. Acoust. Soc. Am. 128, 3211-3222.

Kastelein, R. A., Janssen, M., Verboom, W. C. and de Haan, D. (2005a). Receiving beam patterns in the horizontal plane of a harbor porpoise (*Phocoena phocoena*). J. Acoust. Soc. Am. 118, 1172-1179.

Kastelein, R. A., Steen, N., Gransier, R., Wensveen, P. J. and de Jong, C. A. (2012b). Threshold received sound pressure levels of single 1-2 kHz and 6-7 kHz up-sweeps and down-sweeps causing startle responses in a harbor porpoise (*Phocoena phocoena*). J. Acoust. Soc. Am. 131, 2325-2333.

Kastelein, R. A., Verboom, W. C., Jennings, N. and de Haan, D. (2008a). Behavioral avoidance threshold level of a harbor porpoise (*Phocoena phocoena*) for a continuous 50 kHz pure tone. J. Acoust. Soc. Am. 123, 1858-1861.

Kastelein, R. A., Verboom, W. C., Jennings, N., de Haan, D. and van der Heul, S. (2008b). The influence of 70 and 120 kHz tonal signals on the behavior of harbor porpoises (*Phocoena phocoena*) in a floating pen. Mar. Environ. Res. 66, 319-326.

Kastelein, R. A., Verboom, W. C., Muijsers, M., Jennings, N. V and van der Heul, S. (2005b). The influence of acoustic emissions for underwater data transmission on the behaviour of harbour porpoises (*Phocoena phocoena*) in a floating pen. Mar Env. Res 59, 287-307.

Kastelein, R. A., Wensveen, P. J., Hoek, L., Au, W. W. L., Terhune, J. M. and de Jong, C. A. F. (2009). Critical ratios in harbor porpoises (*Phocoena phocoena*) for tonal signals between 0315 and 150 kHz in random Gaussian white noise. J. Acoust. Soc. Am. 126, 1588-1597.

Kastelein, R. A., Wensveen, P. J., Terhune, J. M. and de Jong, C. A. F. (2011b). Near-threshold equal-loudness contours for harbor seals (*Phoca vitulina*) derived from reaction times during underwater audiometry: A preliminary study. J. Acoust. Soc. Am. 129, 488-495.

Ketten, D. R. (1997). Structure and function in whale ears. Bioacoustics 8, 103-135.

Kinsler, L. E., Frey, A. R., Coppens, A. B. and Sanders, J. V. (2000). Fundamentals of Acoustics (4th edn). Hoboken, NJ: John Wiley & Sons.

Knudsen, V. O., Alford, R. S., and Emling, J. W. (1948). Underwater ambient noise. J. Mar. Res. 7, 410-429.

Kohfeld, D. L., Santee, J. L. and Wallace, N. D. (1981). Loudness and reaction time: I. Percept. Psychophys. 29, 535-549.

Leibold, L. J. and Werner, L. A. (2002). Relationship between intensity and reaction time in normal-hearing infants and adults. Ear Hear. 23, 92-97.

Luce, R. D. (1986). Response times: their role in inferring elementary mental organization. New York, NY: Oxford University Press.

Luce, R. D. and Green, D. M. (1972). A neural timing theory for response times and the psychophysics of intensity. Psychol. Rev. 79, 14-57.

Lucke, K., Siebert, U., Lepper, P. A. and Blanchet, M. A. (2009). Temporary shift in masked hearing thresholds in a harbor porpoise (*Phocoena phocoena*) after exposure to seismic airgun stimuli. J. Acoust. Soc. Am. 125, 4060-4070.

Madsen, P. T. (2005). Marine mammals and noise: problems with root mean square sound pressure levels for transients. J. Acoust. Soc. Am. 117, 3952-3957.

Marks, L. E. and Florentine, M. (2011). Measurement of Loudness, Part I: Methods, Problems, and Pitfalls. In Loudness (ed. M. Florentine, A. N. Popper and R. R. Fay), pp. 17-56. New York, NY: Springer.

Marshall, L. and Brandt, J. F. (1980). The relationship between loudness and reaction time in normal hearing listeners. Acta Otolaryngol. 90, 244-249.

May, B. J., Little, N. and Saylor, S. (2009). Loudness perception in the domestic cat: reaction time estimates of equal loudness contours and recruitment effects. J. Assoc. Res. Otolaryngol. 10, 295-308.

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

Moody, D. B. (1973). Behavioral studies of noise-induced hearing loss in primates: loudness recruitment. Adv. Otorhinolaryngol. 20, 82-101.

Møhl, B. and Andersen, S. (1973). Echolocation: high-frequency component in the click of the harbour porpoise (*Phocoena ph*. L.). J. Acoust. Soc. Am. 54, 1368-1372.

Mulsow, J. and Finneran, J. J. (2013). Auditory reaction time measurements and equal-latency curves in the California sea lion (*Zalophus californianus*) and bottlenose dolphin (*Tursiops truncatus*). Proc. Meet. Acoust. 19, 010003.

Nedwell, J. R., Turnpenny, A. W. H., Lovell, J. M. and Edwards, B. (2006), An investigation into the effects of underwater piling noise on salmonids. J. Acoust. Soc. Am. 120, 2550-2554.

Pfingst, B. E., Hienz, R., Kimm, J. and Miller, J. (1975a). Reaction-time procedure for measurement of hearing. I. Suprathreshold functions. J. Acoust. Soc. Am. 57, 421-430.

Pfingst, B. E., Hienz, R. and Miller, J. (1975b). Reaction-time procedure for measurement of hearing. II. Threshold functions. J. Acoust. Soc. Am. 57, 431-436.

Pins, D. and Bonnet, C. (2000). The Pieron function in the threshold region. Percept. Psychophys. 62, 127-136.

Piéron, H. (1920). Nouvelles recherches sur l'analyse du temps de latence sensorielle et sur la loi qui relie le temps a l'intensité d'excitation. Ann. Psychol. 22, 58-142.

Popov, V. V, Supin, A. Y., Wang, D., Wang, K. X., Dong, L. J. and Wang, S. Y. (2011). Noise-induced temporary threshold shift and recovery in Yangtze finless porpoises *Neophocaena phocaenoides asiaeorientalis*. J. Acoust. Soc. Am. 130, 574-584.

Reason, J. T. (1972). Some correlates of the loudness function. J. Sound Vibration 20, 305-309.

Ridgway, S. H. and Carder, D. A. (2000). A preliminary study of loudness at frequencies of 5 to 120 kHz based on whistle response time (RT) in a dolphin. J. Acoust. Soc. Am. 108, 2515.

Ridgway, S. H., Carder, D. A., Kalmolnick, T., Smith, R. R., Schlundt, C. E. and Elsberry, W. R. (2001). Hearing and whistling in the deep sea: depth influences whistle spectra but does not attenuate hearing by white whales (*Delphinapterus leucas*) (Odontoceti, Cetacea). J. Exp. Biol. 204, 3829-3841.

Sanders, A. F. (1998). Elements of human performance: Reaction processes and attention in human skills. Mahwah, NJ: Lawrence Erlbaum Associates.

Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene Jr., C. R., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., et al. (2007). Marine mammal noise exposure criteria: Initial scientific recommendations. Aquat. Mamm. 33, 411-521.

Stebbins, W. C. (1966). Auditory reaction time and the derivation of equal loudness contours for the monkey. J. Exp. Anal. Behav. 9, 135-142.

Stebbins, W. C. and Miller, J. M. (1964). Reaction time as a function of stimulus intensity for the monkey. J. Exp. Anal. Behav. 7, 309-312.

Stevens, S. S. (1955). The measurement of loudness. J. Acoust. Soc. Am. 27, 815–829.

Suzuki, Y. and Takeshima, H. (2004). Equal loudness-level contours for pure tones. J. Acoust. Soc. Am. 116, 918-933.

Takeshima, H., Suzuki, Y., Ozawa, K., Kumagai, M. and Sone, T. (2003). Comparison of loudness functions suitable for drawing equal loudness-level contours. Acoust. Sci. Technol. 24, 61-68.

Tougaard, J., Wright, A. J. and Madsen, P. T. (2015). Cetacean noise criteria revisited in the light of proposed exposure limits for harbour porpoises. Mar. Pollut. Bull. 90, 196-208.

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

Verboom, W. C. and Kastelein, R. A. (2005). Some examples of marine mammal discomfort thresholds in relation to man-made noise. Proceedings of the Undersea Defence Technology Conference 2005, Amsterdam, the Netherlands.

Villadsgaard, A., Wahlberg, M. and Tougaard, J. (2007). Echolocation signals of wild harbour porpoises, *Phocoena phocoena*. J. Exp. Biol. 210, 56-64.

Wagner, E., Florentine, M., Buus, S. and McCormack, J. (2004). Spectral loudness summation and simple reaction time. J. Acoust. Soc. Am. 116, 1681-1686.

TABLES & FIGURES

Table 4.1. The root-mean-square error (RMSE) between the SPLs of the unsmoothed equal latency contour and the SPLs of the smoothed equal latency contour, as functions of the test frequencies included. The highest similarity between the two curves, indicated by the smallest RMSE (underlined), was generally obtained with the 0.5−31.5 kHz data set.

Contour	**Reference RT (ms)**	**Frequencies included (kHz)**				
		≤16	**≤31.5**	**≤63**	**≤80**	**≤125**
		RMSE (dB)				
I	150	4.6	4.3	7.2	10.4	10.7
II	160	3.4	3.2	6.0	8.3	8.2
III	170	3.1	3.0	5.2	6.6	6.3
IV	180	3.8	3.7	5.0	5.5	5.2
V	190	5.0	4.7	5.3	5.1	5.3
VI	200	6.4	5.9	5.9	5.5	6.2

Table 4.2. Summary of the reaction time (RT) data, and fitted auditory RT functions (Eqn 4.1), including the total number of trials, the total number of test levels, the best-fit estimates and 95% confidence intervals (CIs) of parameters α and β (Eqn 4.1), and the goodness-of-fit parameters: r^2 and RMSE. Also shown are the hearing thresholds of the same porpoise for the test frequencies in this study (900 ms signals; Kastelein et al., 2010), which were used for the conversion between SPL and SnL.

Frequency (kHz)	**No. of trials**	**No. of test levels**	**Log-log slope α**		**Intercept β (ms)**		**Goodness of fit**		**Hearing threshold (dB re 1 µPa)**
			Best-fit value	95% CI	Best-fit value	95% CI	r^2	RMSE (ms)	
0.5	493	7	0.042	0.006	278.1	15.0	0.99	3.9	102
1	598	8	0.028	0.004	243.9	11.8	0.98	3.3	85
2	498	8	0.035	0.006	260.5	19.4	0.98	4.1	71
4	756	10	0.032	0.004	270.7	13.0	0.98	5.1	60
16	537	8	0.019	0.007	217.7	18.0	0.90	6.0	49
31.5	578	9	0.022	0.006	234.7	19.5	0.93	6.4	47
63	527	8	0.030	0.005	234.9	13.2	0.98	4.4	55
80	500	8	0.045	0.004	265.3	12.6	0.99	2.7	46
125	657	9	0.056	0.009	327.9	42.5	0.98	5.9	43

Table 4.3. Information on the captive studies with four harbour porpoises from which behavioural response thresholds were obtained.

Ref.	**Sound characteristics**							**Behavioural response**		
	Peak freq. (Hz)	Band-width	Signal type	Pulse dur. (s)	Duty cycle (%)	Corr. to $L_{eq\text{-}fast}$ (dB) [c]	SPL at 1 m (dB re 1 μPa)	$L_{eq\text{-}fast}$ (dB re 1 μPa)	No. of anim.	Comment
Kastelein et al. (2005b)	12	10-13 kHz, with harmonics	Chirp	2	80	0	116	98[d]	2	Avoidance / discomfort threshold was assumed to occur just outside of the average swimming area. Thresholds for three porpoises in total; not the same as in the present study.
	12	9-18 kHz, with harmonics	Direct seq. spread spectrum	1	60	0	123	106[d]	2	
	12	10-14 kHz	Linear upsweep	1	80	0	116	98[d]	2	
	12	10-14 kHz	Mod.freq.shift keying (noise)	Cont.	100	0	130	112	2	
Kastelein et al. (2008a)	50	Narrow	Pure tone	Cont.	100	0	122	108	1	
Kastelein et al. (2008b)	70	Narrow	Pure tone	0.3	8	–0.4	137	122[e]	2	
	70	Narrow	Pure tone	1	25	0	137	122[e]	2	
	70	Narrow	Pure tone	Cont.	100	0	148	137[e]	2	
	70	Narrow	Pure tone	2	50	0	144	127	2	
	120	Narrow	Pure tone	Cont.	100	0	134	117[e]	2	
Kastelein et al. (2012b)	1[a]	1-2 kHz	Hyperbolic upsweep	1	0.6	0	132	132	1	Response thresholds taken as the 50%
	2[a]	1-2 kHz	Hyperbolic	1	0.6	0	132	134	1	

			downsweep							probability of a brief change in speed and/or orientation. Thresholds for the same porpoise as in the present study.
	6[a]	6-7 kHz	Hyperbolic upsweep	1	0.6	0	114	101	1	
	7[a]	6-7 kHz	Hyperbolic downsweep	1	0.6	0	114	101	1	
	12[b]	1-2 kHz, with harmonics	Hyperbolic upsweep	1	0.6	0	110	98	1	
	12[b]	1-2 kHz, with harmonics	Hyperbolic downsweep	1	0.6	0	110	100	1	
Kastelein et al. (2013b)	1.38	Narrow	Pure tone	1.25	0.7	0	-	131	1	
	1.38	Narrow, with harmonics	Pure tone	1.25	0.7	0	-	124	1	
	1.38	Narrow, with harmonics	Pure tone, tapered	1.25	0.7	0	-	125	1	
	1.43[a]	1.33-1.43 kHz	Hyperbolic downsweep	1.25	0.7	0	-	144	1	
	1.43[a]	1.33-1.43 kHz, with harmonics	Hyperbolic downsweep	1.25	0.7	0	-	130	1	
Kastelein et al. (2015a)	25	24.5-25.5 kHz, with harmonics	Hyperbolic downsweep	0.05	2.4	–5	160	120	1	Response thresholds taken as the lowest levels where respiration rate and/or swimming speed was
	25	Narrow, with harmonics	Amplitude modulated tone, bell shaped	0.5	5.6	0	160	118	1	
	25	24.5-25.5	Combo of the two	0.9	8.3	0	159	118	1	

		kHz, with harmonics	above							increased. Thresholds for the same porpoise as in the present study.

a) The start frequency of the sweep was used because the observed behavioural response was assumed to be the result of a startle reflex

b) Responses were assumed to be triggered by the harmonic energy around 12 kHz (based on sensation level; Kastelein et al., 2011a)

c) The correction to $L_{eq,fast}$ was calculated as $10 * \log_{10}(1 - e^{-t/0.125})$ with signal duration t in seconds (Tougaard et al., 2015)

d) The influence of silences between pulses on the reported levels was corrected by subtracting $10 * \log_{10}(d / 100)$ with duty cycle d in %

e) Avoidance threshold was estimated by adding the difference in "SPL in area occupied by porpoises" to the reported avoidance threshold for the 2-s signal

Table 4.4. Parameter estimates with 95% confidence intervals (CIs) for the hypothetical equal loudness contours (Eqn 3; Fig. 4.7B) and the auditory weighting functions (Eqn 4.4; Fig. 4.8). Note that the 95% CIs are the same for parameters K_1 and K_2 because K_2 was derived from K_1. The low frequency roll-off rate, which depends on parameter *x*, is also shown.

Contour	**Parameter**								**Roll-off rate (dB/decade)**
	K_1 (dB re 1 μPa)		K_2 (dB)		*a* (Hz)		x		
	Est.	95% CI	Est.	95% CI	Est.	95% CI	Est.	95% CI	
I	133.48	3.32	0.03	3.32	4563	3322	1.62	0.58	32
II	121.17	3.80	0.02	3.80	5004	4180	1.84	0.65	37
III	109.62	4.27	0.01	4.27	5318	4393	2.05	0.71	41
IV	98.71	4.73	0.01	4.73	5550	4550	2.26	0.78	45
V	88.41	5.17	0.00	5.17	5722	4667	2.45	0.84	49
VI	78.62	5.60	0.00	5.60	5856	4758	2.64	0.9	51

Table 4.5. Results of fitting the vertical offset of each idealised equal loudness contours to the 1) TTS onset data for frequency sweeps, 2) TTS onset data for frequency sweeps and octave-band nose, and 3) behavioural response onset thresholds. Maximum likelihood estimates for parameter K_1 and measures of goodness-of-fit are shown for each data set.

Contour	**TTS onset (sweeps)**			**TTS onset (sweeps & noise)**			**Behavioural response**		
	K_1	r^2	RMSE	K_1	r^2	RMSE	K_1	r^2	RMSE
I	177.3	0.88	2.46	171.5	0.40	9.48	107.4	0.73	10.13
II	175.5	0.77	3.42	170.0	0.43	9.31	105.5	0.80	8.75
III	173.9	0.55	4.73	168.7	0.42	9.33	103.7	0.84	7.78
IV	172.3	0.24	6.17	167.3	0.40	9.53	102.0	0.86	7.28
V	170.9	-0.13	7.54	166.1	0.35	9.87	100.4	0.86	7.29
VI	169.5	-0.59	8.94	164.9	0.29	10.33	98.9	0.84	7.75

Figure 4.1. The indoor test pool, as seen (A) from above and (B) from the side. The harbour porpoise, trainer, and signal operator are shown in their positions at the beginning of a research trial. Signals were transmitted with a low frequency (LF; 0.5−4 kHz), mid frequency (MF; 16−63 kHz), or high frequency (HF; 80−125 kHz) projector.

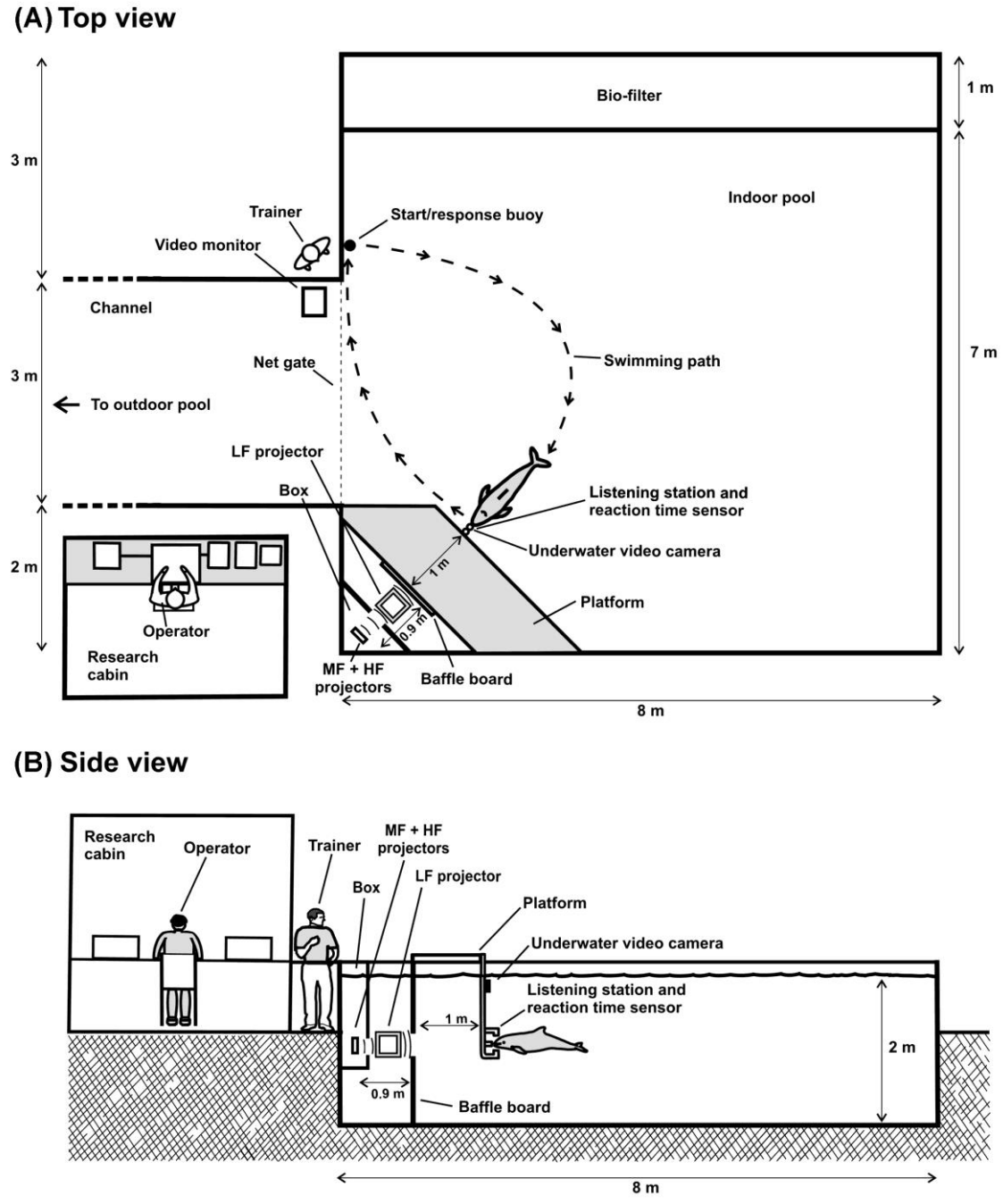


Figure 4.2. Block diagram of the equipment used to generate the sound stimuli, monitor the sounds and the harbour porpoise under water, and measure the porpoise's RTs.

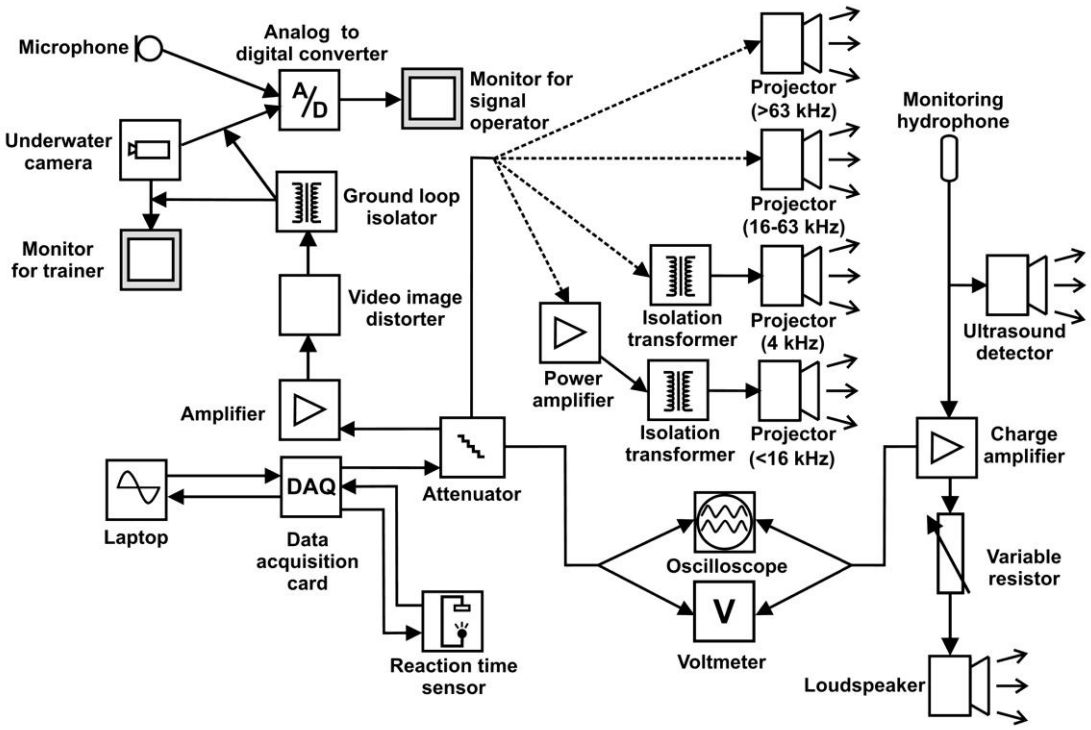


Figure 4.3. The listening station with embedded reaction time sensor based on IR light.

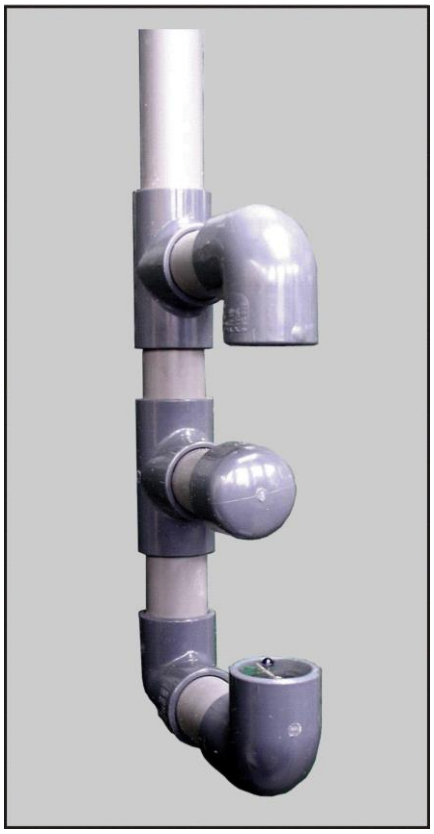


Figure 4.4. Still frame sequence of a research trial. (A) The harbour porpoise approaches the listening station. (B) The porpoise positions its rostrum against the listening station, thereby breaking the infrared light beam in the reaction time sensor, and illuminating the LED in the top sensor bracket above the listening station (indicated by the arrow). (C) The porpoise has responded and starts moving away while the sound is still being emitted (horizontal lines produced by the video distorter are visible in the image because the sound is on). (D) The porpoise swims back to the start/response buoy to receive a fish reward from the trainer.

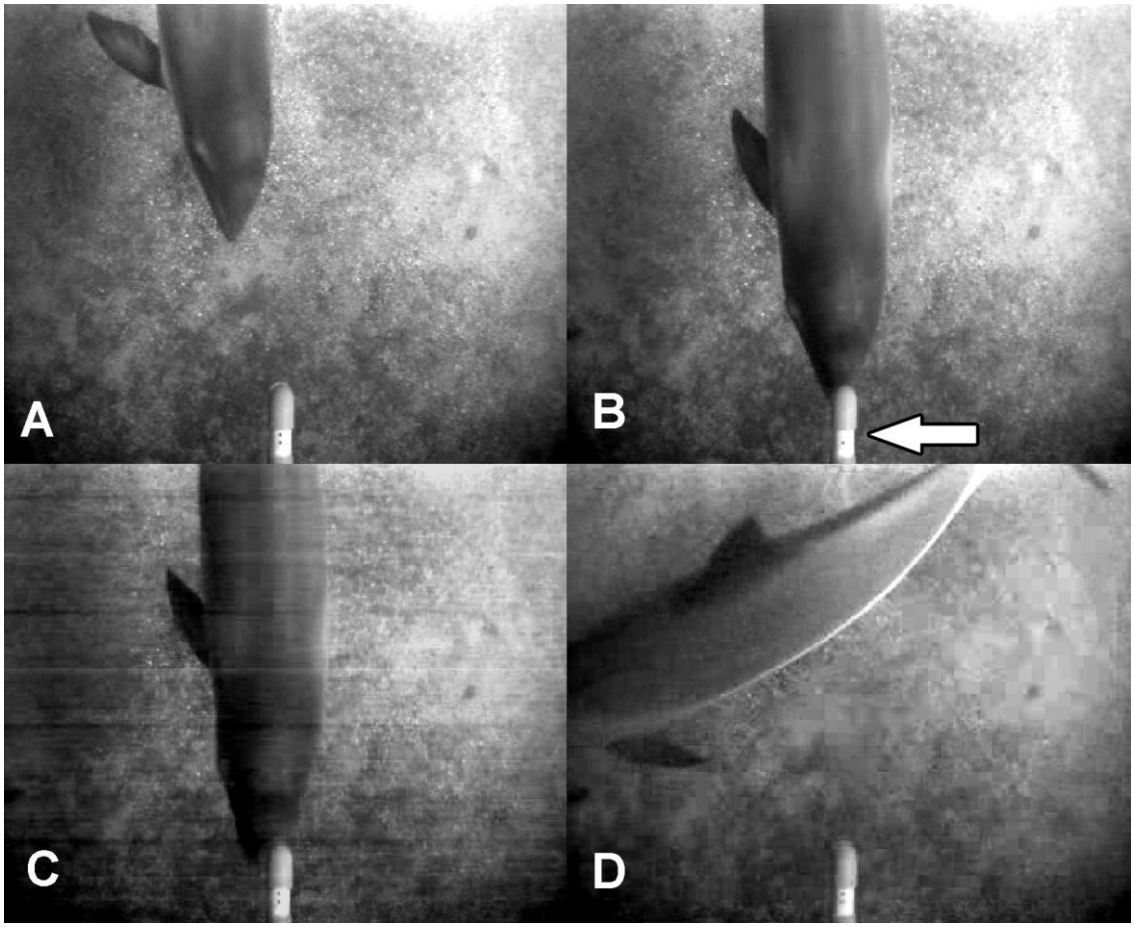


Figure 4.5. Relationship between sensation level (SnL; top axes), sound pressure level (SPL; bottom axes) and reaction time (RT) of the harbour porpoise, for narrowband FM tonal signals with centre frequencies of 0.5 to 125 kHz. The auditory RT functions (black lines; Eqn 4.1) result from fitting a power law (Eqn 4.1) to the median RTs (circles), after one or two near-threshold medians (crosses) were omitted. The number of RT measurements is shown above each median, and error bars indicate interquartile ranges.

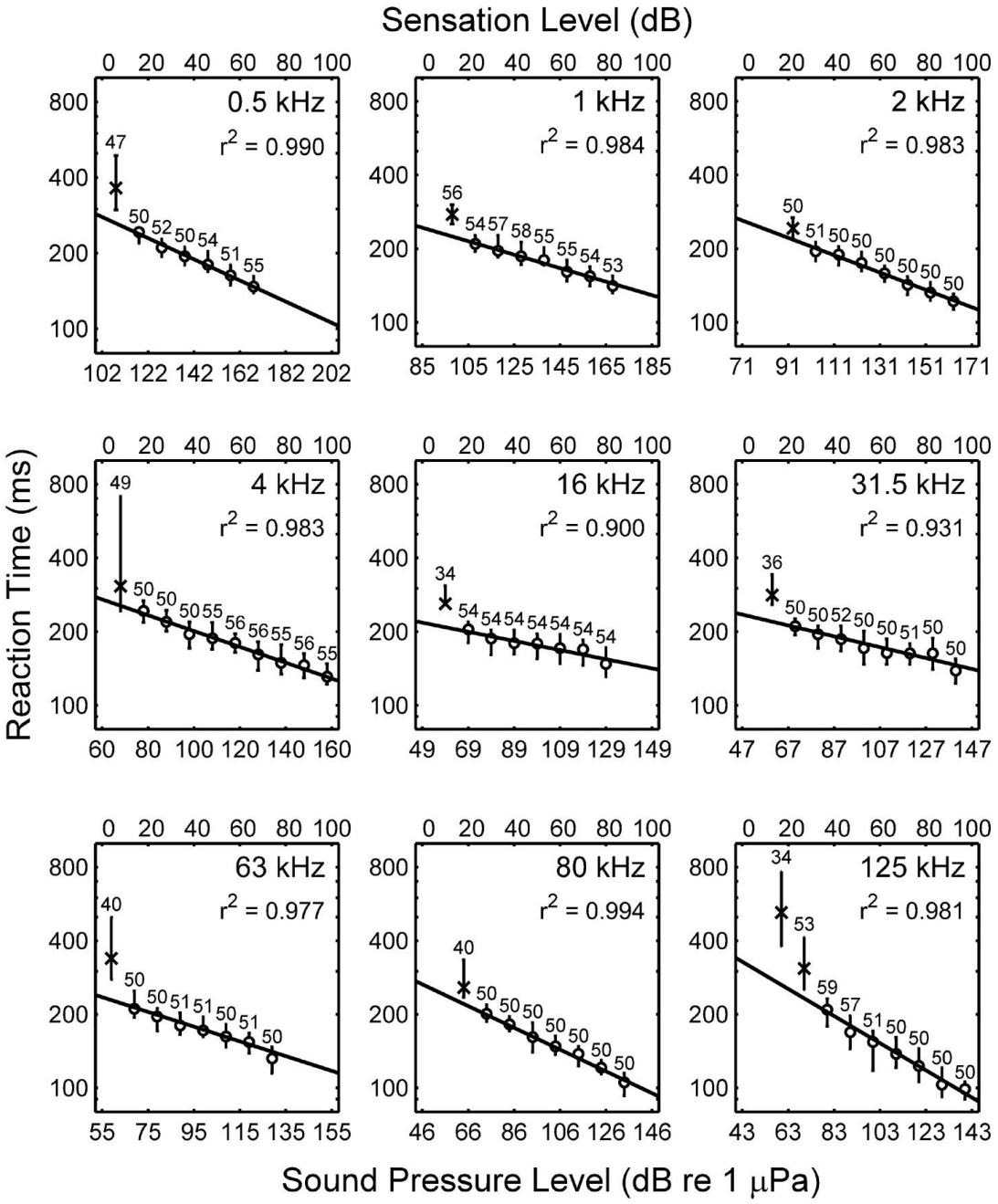


Figure 4.6. Equal latency contours I−VI of the harbour porpoise with the corresponding reference RT values. The contour SPLs were derived from the auditory RT functions in Fig. 4.5 by matching the six reference RTs across frequencies. Circles and crosses indicate the test levels for which median RTs were included and excluded during the fitting process, respectively (see 'Materials and Methods – Reaction time measurements' for rationale). The line with dotted markers at the bottom is the porpoise's hearing threshold for 900 ms signals (Kastelein et al., 2010).

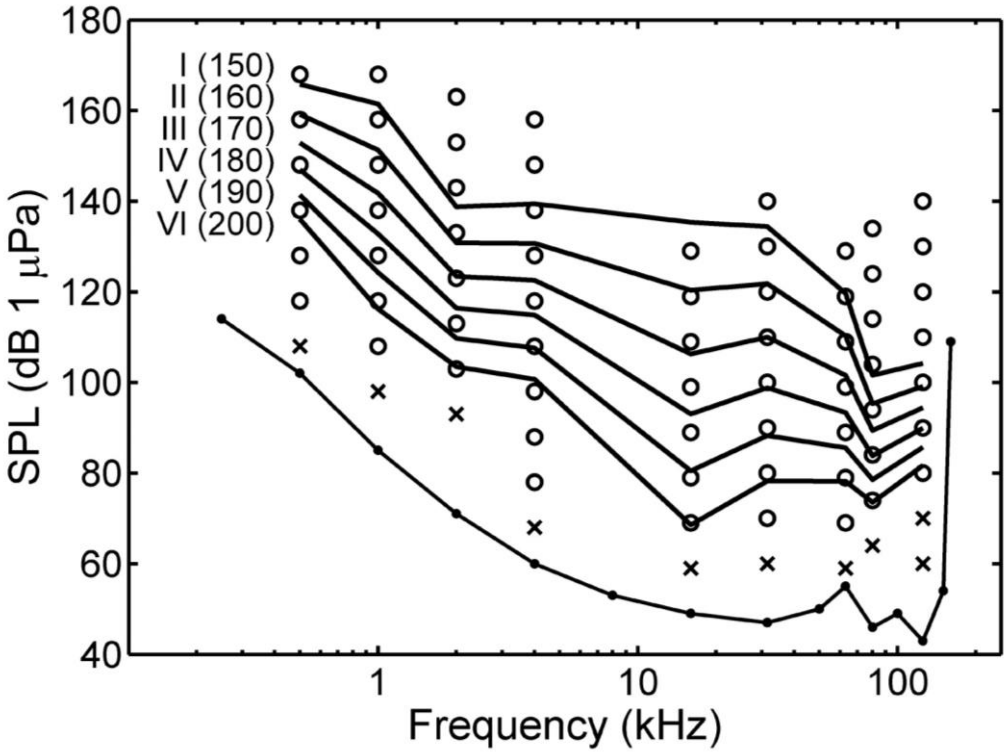


Figure 4.7. Conversion from equal latency contours to equal loudness contours. (A) The shape of the harbour porpoise's audiogram was used as a template to create smoothed versions of the equal latency contours (solid lines with squares) from the original contours (dashed lines; same as in Fig. 4.5). The audiogram of the subject (solid line with dots; Kastelein et al., 2010) and the audiograms of two other harbour porpoises (dashed-dotted line: Andersen, 1970; dotted line: Kastelein et al., 2010) are shown at the bottom of the graph. (B) Values for 0.25, 8, and 50−150 kHz were added to the smoothed equal latency contours using the animal's own hearing thresholds at these frequencies and the threshold-contour relationships (Eqn 4.2), which resulted in six hypothetical equal loudness contours (dashed lines with squares). A closed-form model (Eqn 4.3) was fitted (solid lines) to these extended smoothed contours. Only the audiogram of the subject is shown here (solid line with dots).

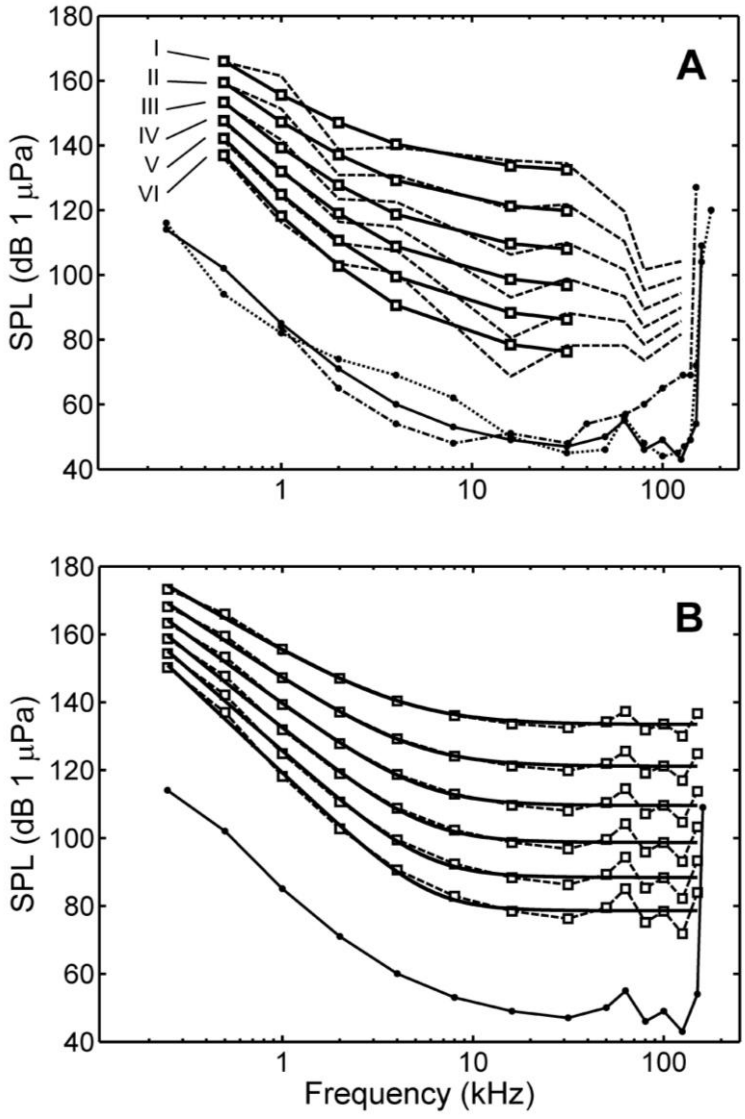


Figure 4.8. Six auditory weighting functions for the harbour porpoise. The weighting functions, which should reflect the frequency response of the porpoise's hearing system, are associated with the SnL of the received signal (I being high and VI being low). The extension of the functions (dashed line) is to emphasize that the effective hearing range of the porpoise ends abruptly at 160 kHz.

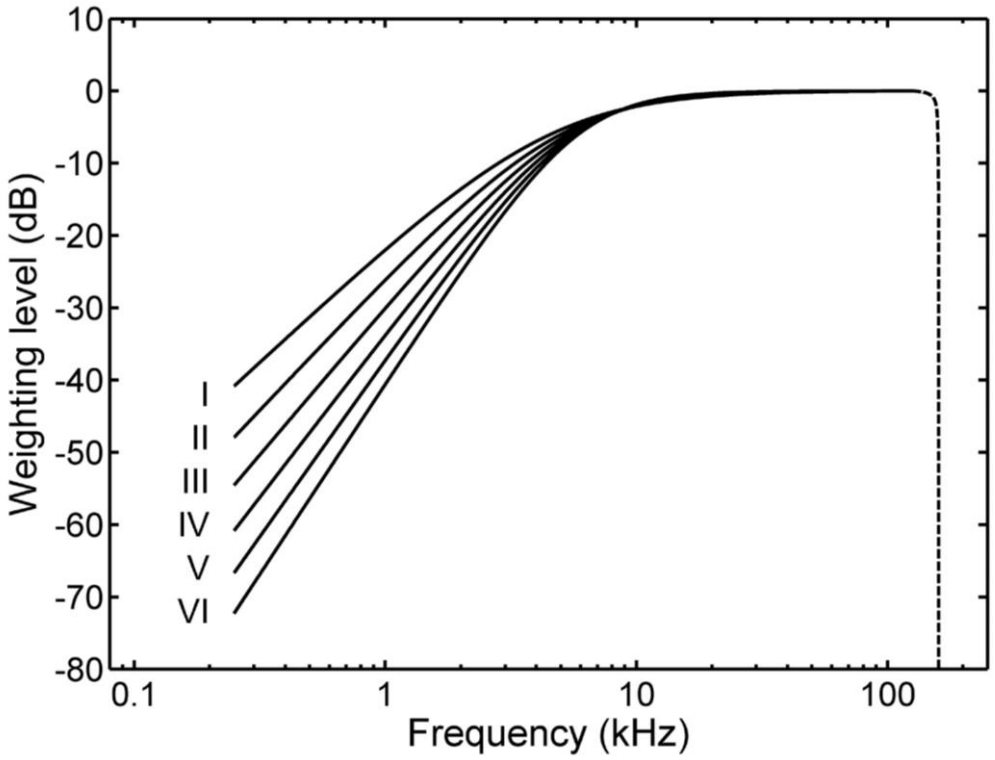
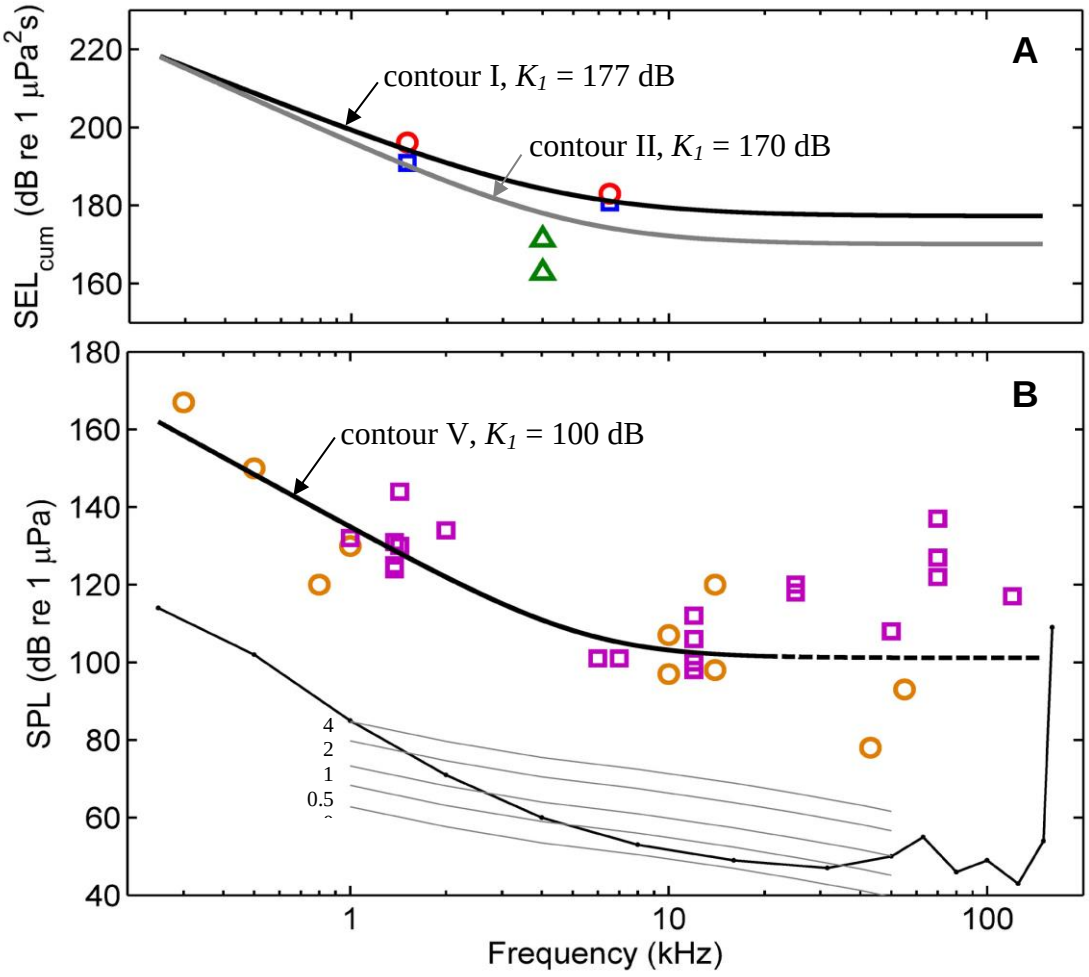


Figure 4.9. (A) TTS onset thresholds of porpoise Jerry measured after exposure to downsweeps of 1–2 kHz (Kastelein et al., 2014) and 6–7 kHz (Kastelein et al., 2015b), for signal sequences with duty cycles of 10% (circles) and 100% (squares). This data set was best described by the shape of contour I (black line). The shape of contour II (grey line) best described the TTS onset thresholds if the data set also included 4 kHz octave-band noise exposures (triangles; Kastelein et al., 2012a). (B) Behavioural response thresholds derived from field studies (circles; reviewed by Tougaard et al., 2015) and captive studies (squares; see Table 4.3 for references). The shape of contour V best predicted the behavioural response thresholds for frequencies below 20 kHz. Thresholds for higher frequencies are also shown but were not included in the final model. The audiogram of the porpoise and the noise-limited theoretical detection thresholds for 5 different sea states are shown at the bottom of the panel.



Chapter 5

Experimental evaluation of the effectiveness of ramp-up as a mitigation method for naval sonar

SUMMARY

Powerful sound sources such as naval sonars are often gradually increased in intensity to reduce risk of physiological damage, by giving animals an opportunity to move away and reduce their exposure to harmful levels, but the effectiveness of this 'ramp-up' procedure has yet to be empirically evaluated. Among 31 vessel approaches to 13 humpback whales tagged with multi-sensor tags, ramp-up did not significantly reduce sound exposure overall. However, for the 60% of whales that initially avoided the sonar during ramp-up, statistically significant reductions in received maximum sound pressure level (SPL_{max}: −6 dB) and cumulative sound exposure level (SEL_{cum}: −4 dB) were observed, compared to whales that did not avoid. During a consecutive ramp-up, fewer avoidance responses (38%) led to smaller reductions in received levels (SPL_{max}: −2; SEL_{cum}: 0 dB), indicating short-term habituation. Attraction responses occurred occasionally (13%), which resulted in small increases in received levels. Strong avoidance without habituation was observed for a whale accompanied by a small calf. This chapter demonstrated that 1) sonar ramp-up can reduce received levels and thereby risk of acute effects in humpback whales, and 2) ramp-up is only effective for the population subset that shows strong enough avoidance at sufficiently low received levels.

INTRODUCTION

The importance of sound for marine animals has led to increased recognition that man-made noise is a marine pollutant (Tasker et al., 2010) and that exposure to man-made noise can be detrimental to marine mammals which have acute underwater hearing (Richardson et al., 1995). Best-practise guidelines for human activities that generate

high-intensity underwater sound currently recommend the use of operational mitigation measures designed to protect marine animals (e.g. Compton et al., 2008).

One such 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 sound sources, e.g. seismic surveys for oil and gas exploration (Nowacek et al., 2013), acoustic thermometry (Frankel and Clark, 2000a), and pile driving and detonation of explosions during offshore construction (Brandt et al., 2011; dos Santos et al., 2010; Jefferson et al., 2009). The untested assumption behind ramp-up is that animals will move away from the path of an approaching sound source or the location of a stationary source, while the source level (SL) increases, thus reducing the maximum sound intensity and energy received by the animal. A recent non-peer-reviewed correlational study based on observational data collected over 7 years during seismic surveys provided evidence that ramp-up indeed triggers such avoidance behaviour in cetaceans (Stone, 2015). Ramp-ups are mostly intended to mitigate against auditory (Mooney et al., 2009) and other types of physiological damage in animals that are relative 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.

von Benda-Beckmann et al. (2014) used theoretical modelling to investigate the effectiveness of ramp-up for naval sonar. These authors found that ramp-up before normal sonar operation can be effective at reducing the number of animals experiencing sound doses that are high enough to cause temporary or permanent hearing loss. Important factors were the assumed relationship between acoustic dosage and avoidance response speed of the animals, 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 von Benda-Beckman et al. (2014) used relevant empirical input in their model, ultimate experimental confirmation of these predictions is still missing. Some argue that ramp-up affects the fidelity of naval combat training, and in theory ramp-up might do more harm than good by attracting animals to the source when levels are low, or by increasing the total duration and energy of exposure (Barlow and Gisiner, 2006). It is

therefore important to address the effectiveness of ramp-up as a procedure to mitigate risk to marine mammals during sonar operations.

Experiments in which controlled doses of a stimulus are applied to induce a behavioural or physiological response have been used to study the responses of cetaceans to naval sonar. Such controlled exposure experiments (CEEs; Tyack et al., 2003) have provided useful information about various aspects of the behavioural responses of cetaceans to sonar (DeRuiter et al., 2013; Goldbogen et al., 2013; Kastelein et al., 2011, 2012; Kvadsheim et al., 2012b; Miller et al., 2012, 2015; Sivle et al., 2015; Stimpert et al., 2014; Tyack, 2009; Tyack et al., 2011; Wensveen et al., 2015), including species-specific dose-response relationships (Antunes et al., 2014; Harris et al., 2015; Houser et al., 2013a, 2013b; Miller et al., 2014; Chapter 3) that can be used in environmental risk assessment frameworks (Boyd et al., 2008; Tyack et al., 2003). In this thesis chapter, the objective was to experimentally test the effectiveness of ramp-up of naval sonar in a cosmopolitan cetacean species, the humpback whale (*Megaptera novaeangliae*). The overarching goal of the experiments was to increase understanding about the biological factors that influence the effectiveness of ramp-up in marine mammals in general. Data from CEEs that were specially designed to test whether ramp-up of 1.3-2 kHz active sonar reduces sound levels received by the humpback whales were analysed. The results are discussed in the broader context of marine mammal behaviour and operational sonar use.

MATERIALS & METHODS

In this chapter, I used data from experiments conducted on tagged humpback whales in 2011 and 2012 during the 3S Behavioural Response Study (Kvadsheim et al., 2015; Sivle et al., 2015b). Whales were tagged with Fastloc-GPS loggers (F2G 134A, Sirtrack, New Zealand) and multi-sensor DTAGs (version 2; Johnson and Tyack, 2003) prior to the sonar exposure sessions and whales were focally followed throughout the period that the tag was attached. Readers are referred to Chapter 2 for details about the data loggers and focal follow protocols. Detailed information about the study area and CEE protocols can be found in Chapter 3. This information is summarised here, but I also provide some extra information specific to the current study.

CEE protocol

The source vessel was the 55-m R/V HU Sverdrup II, which towed a sonar source (Socrates II, TNO, The Netherlands) using 250-300 m of cable. The source was positioned at 50-m depth. Three types of exposure session were conducted: *No-Sonar* (n=11), *RampUp* (n=18), and *No-RampUp* (n=2) (Table 3.1). During *RampUp*, sonar transmission started at 1.3 km from the whale, and the source vessel approached at a speed of 4 m/s on a predetermined straight intercept course, while transmitting a 5-minute ramp-up. Full-power transmissions then continued for another 5 minutes. Each exposure session was conducted using two independent intercept calculators to advise the experimental coordinator because the start time of the session, 5 minutes before the estimated intercept with the whale, and line-of-approach depended upon a prediction of the future track of the whale (Kvadsheim et al., 2011).

The two intercept calculator tools were MARIA, developed for the Norwegian Navy by FFI, and RU-tool, custom-built in MATLAB by the author. Both tools used the GPS positions of the source vessel and estimated locations of whale to calculate the time and position of the intercept and the best sailing path of the source ship. The two tools used somewhat different logic: MARIA primarily used the automatic identification system (AIS) signal from the tracking boat, which was usually within 100 m of the whale; and RU-tool predicted the future movements of the whale from the last few sighting positions by taking into account the whale's general movement pattern (i.e. nondirectional, directional without bias in turn angle, or directional with bias in turn angle). These sighting positions were relayed via VHF radio by the observers on the tracking boat. The final course correction and start time of the session (i.e. 5 minutes before the calculated intercept point with the source) was determined only minutes before the start of the session. Ultimate decisions on course changes and start of transmission were always made by the experimental coordinator based on all available information.

To contrast the ramp-up approaches, approaches without ramp-up were also conducted to estimate the received levels animals would experience that were not warned by the

ramp-up and to control for the potential effect of the vessel. These session types followed the same navigational protocol as *RampUp*, but had no transmissions during the first 5 minutes and then started at full power (*No-RampUp*) or no transmissions during the entire 10-minute period (*No-Sonar*) (Table 3.1). Most *No-RampUp* sessions used in the analysis were simulated by prediction of received levels based upon geometry and timing of *No-Sonar* sessions. Specifically, sonar transmissions were modelled using the measurements of the whale depth from the DTAG, source depth, source distance and acoustic environment, thus predicting what the received level at the whale would have been if the *No-Sonar* was a *No-RampUp* session. This approach was taken because of the *No-RampUp* protocol did not give whales enough time to affect received levels by moving away, and the first transmissions always caused the highest received levels. However, two experimental *No-RampUp* sessions were conducted in 2012 (Table 3.1) to investigate potential effects of the sudden nearby sonar onset on the magnitude of the response. Up to three sessions, each separated by 1 hour or more, were performed per experiment, with *No-Sonar* always conducted before sonar exposures to avoid sensitisation to the sound stimulus.

The sonar source transmitted a 1.3-2 kHz hyperbolic upsweep at a 20-s interval. During the ramp-up period, the single-pulse source level based on mean-square pressure (SL) increased over 5 minutes as follows: 152, 168, 180, 187, 192, 196, 199, 202, 204, 206, 208, 210, 211, 213 and 214 dB re 1µPa m. The SL was 214 dB re 1µPa m during the full-power period. Sonar pulses were shorter during ramp-up (0.5 s) than during full-power periods (1 s) to minimise SEL_{cum}. See the Discussion for how the ramp-up scheme was selected and for the rationale of using two different pulse durations.

Acoustic data analysis

Acoustic propagation loss (PL) was modelled at each exposure site using ray-trace software (BELLHOP, version 09/2010; (Porter and Bucker, 1987)) to calculate the received levels during *RampUp* and *No-RampUp* sessions. Sound speed profiles used in the analysis were collected on-site using a conductivity-temperature-depth (CTD) profiler or an expendable bathythermograph, during the exposure session or shortly after the session had ended (Kvadsheim et al., 2011, 2012). 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 from (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 exposure session, incoherent PL was modelled for a single two-dimensional slice with 1×1 m grid resolution. Each slice was 4 km long, and its vertical dimension was taken as the mean sea floor depth between the source and whale at the start of the exposure session (British Oceanographic Data Centre, 2010).

The modelled sound source was based on the properties of the real sonar source. PL was modelled at 1.6 kHz, which was the logarithmic middle of the band. The vertical directivity pattern of the real source at that frequency (3-dB beamwidth: ±40°) was implemented. The range of beam take-off angles in the vertical plane was ±89°. 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 exposure session.

The energy source level (SL_E) was derived from the SL via $SL_E=SL+10\log_{10}(T/t_{ref})$, where t_{ref} is 1 s. Because of a gradual onset and offset in the waveform, the effective duration T of a pulse transmitted during the ramp-up and full-power periods was 0.43 and 0.93 s, respectively. The received single-pulse sound exposure level (SEL) and sound pressure level (SPL) were derived from the modelled PL as $SEL=SL_E–PL$ and $SPL=SL–PL$, respectively [6]. Received SEL_{cum} for a given 3-dimensional whale trajectory was calculated by cumulative summation of the single-pulse sound exposures E since the start of the session, where $E=10^{SEL/10}$. Received SPL_{max} was the maximum of all the received single-pulse SPLs in the session.

Instead of using a deterministic approach with point estimates of PL, a Monte Carlo method was applied to propagate forward uncertainty of the whale position into probabilistic received levels (SPL_{max} and SEL_{cum}). For each modelled sonar pulse, a small degree of noise (SD: 1 m) was added to the whale depth measured by the DTAG.

This, in combination with probabilistic source-whale ranges derived from the horizontal tracks (Chapter 2), gave rise to uncertainty in the PLs, which in turn resulted in the probabilistic received levels. Statistical analyses using generalised estimation equations (GEEs) were conducted on the mean values of these probabilistic received levels.

For all humpback whales exposed to sonar, Sivle et al. (2015) describes how the received levels of the pulses were calculated from the acoustic recordings made by the DTAGs. As a performance check of the modelling approach, the measured SELs were compared to the modelled SELs (Fig. 5.1). Good correspondence was found in general, but there was a tendency for the modelled SEL_{cum} at the end of the exposure session to be slightly lower than the measured SEL_{cum} (mean difference 2.2 dB). This value fell roughly within 1 standard deviation (SD) from the mean acoustic sensitivity of the DTAGs in the frequency band of the sonar (SD 1.5-2.6 dB; n=3 tags; Appendix III). Althought it would have been possible to add corrections to the modelled levels to bring them closer to the measured levels, this approach was not taken as it would not affect the main objective, i.e. to study the relative effects of the exposure session type and other covariates on the received levels and mimimum distance.

Qualitative analysis of behaviour

Visual estimates of range and bearing between the observer and one whale in each group (the 'focal animal') were collected. In addition, GPS positions were recorded by the Fastloc-GPS loggers attached to most DTAGs (12/13). Probabilistic horizontal tracks of the whales were reconstructed in Chapter 2 of this thesis from these data in combination with measurements of speed-through-water and body orientation. Here, I qualitatively assigned the presence/absence of avoidance and/or attraction based on these fine-scale movement tracks (Fig. 5.2; for plots of all sessions, see Figs. A14-A24 in Appendix II). Two expert groups of 2 persons each independently scored the behavioural changes, which were thought to be responses to the sonar source or source vessel, and then reached consensus in a joint meeting (for more details on this analysis, see Chapter 3).

Both expert groups had scored the same 9 avoidance responses (100% agreement), and the independent scores for attraction agreed by 75% (3/4 responses) before consensus. All but two of the avoidance responses were also identified in an earlier study with the same whales (Sivle et al., 2015b) that used horizontal tracks of lower temporal resolution (i.e. only surface positions). One of these two avoidance responses (mn12_179a, session #2; Fig. A23) that was not identified in that earlier study occurred entirely underwater, and was therefore difficult to detect without the fine-scale movement tracks.

I specified the presence/absence of feeding just before or during the exposure session based upon the detection of lunges (Figs. 5.2A,C), 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; Simon et al., 2012). 'Feeding' was assigned if at least one lunge was present in the time interval from 10 minutes before the start of the session until either the onset of the avoidance or attraction response, or the end of the session if no response was scored. The feeding lunges that were used here were identified by Sivle et al. (2015) using a published detection algorithm (Simon et al., 2012).

Statistical procedure

GEEs (Hardin and Hilbe, 2003) were used to model three response variables: SPL_{max}, SEL_{cum} and the minimum source-whale range (R_{min}). SPL_{max} and SEL_{cum} were modelled as Gaussian variables; R_{min} was modelled as a Gamma variable. All predictor variables were binary factor covariates. To examine the effect of ramp-up on the humpback whale population as a whole, I first ran simple GEE models with two predictor variables, session type (0:*No-RampUp*, 1:*RampUp*) and ramp-up session order (0:*RampUp1*, 1:*RampUp2*) for each response variable. However, because the movement tracks reflected heterogeneity in the whales' responsiveness, I also conducted stepwise covariate selection on extended models that included 4 additional covariates: the presence/absence of an attraction response, presence/absence of an avoidance response, order of the ramp-up session in which the avoidance occurred (0:*RampUp1*, 1:*RampUp2*), and presence/absence of feeding behaviour. These 6 respective candidate covariates were labelled SessionType, SessionOrder, Attraction, Avoidance,

AvoidanceOrder, and Feeding. I expected to see significant effects for the covariates Attraction, Avoidance, AvoidanceOrder and Feeding if the behavioural state or behavioural responses of the whales strongly affected the received levels or R_{min}. Additional significant effects of the covariates SessionType and SessionOrder were expected in the case of systematic differences that were predominantly not defined by whale behaviour (e.g. timing of the sonar, course of the source vessel).

GEEs allow for the specification of a blocking unit within which observations can be correlated (Hardin and Hilbe, 2003). Here, group ID was selected as the blocking unit because data from both focal whale mn12_170b and non-focal whale mn12_170a were included, and the behaviour of these associated whales may be not independent. Another non-focal whale (mn11_165f; Fig. 5.3B) was not part of the GEE analysis as visual or GPS fixes were not recorded for this animal. The jackknife variance estimator was applied because the sandwich variance estimator can be biased for small sample sizes. Statistical analyses were performed using geepack (Højsgaard et al., 2014) and MuMIn (Barton, 2015) in R version 3.0.2.

Hypothesis-based stepwise model selection using p-values and backwards selection was used to test which combination of predictor variables best explained the observed variation in the extended models. For each response variable, I first ran the full model with all candidate covariates including the six main effect terms and one 2-way interaction term of interest (Feeding:Avoidance). An ANOVA was then conducted on the candidate model, the covariate or interaction term with the highest p-value removed, and the GEE model rerun. This process was repeated until all terms retained in the ANOVA were significant at 5% level (Table 5.1).

An independent correlation structure was used in all GEE analyses. I verified that three competing working correlation structures (i.e. exchangeable, AR1, unstructured) did not improve the QICr (Hardin and Hilbe, 2003) of the simple models and most-parsimonious extended models. The significance of the factor level coefficients was assessed for the simple models and most-parsimonious models. Although the covariate

SessionType was retained in the most-parsimonious model of R_{min}, no significant difference between its factor levels was found so this parameter was not interpreted.

RESULTS

The 13 tagged humpback whales were part of 11 independent groups of 1 or 2 whales; on two occasions two whales in the same group were tagged (mn11_165e and f; mn12_170a and b) (Table 3.1). All 11 independent groups were first subjected to a *No-Sonar* control session. Of these 11, 10 groups were subsequently exposed to *RampUp1*, as one tag came off early (mn11_158a). Then, 8 groups were exposed to *RampUp2* and 2 groups were exposed to *No-RampUp* (Table 3.1).

Overall, *RampUp* did not significantly reduce the predicted received levels in this humpback whale population, although received levels tended to be slightly lower during *RampUp* than during *No-RampUp* (Table 5.2). Of the 10 whales that were exposed to *RampUp1,* 6 avoided the sound source (Fig. 5.4), there were statistically significant decreases in the received levels (−6 dB for SPL_{max}; −4 dB for SEL_{cum}) of these whales that showed avoidance compared to animals that did not show avoidance (Table 5.2). Fewer (3/8 whales) and weaker avoidance responses occurred during *RampUp2*, leading to smaller average decreases in received levels (−2 dB for SPL_{max}; 0 dB for SEL_{cum}) (Table 5.2).

Avoidance was a response to the sonar and not the approaching vessel alone, as this response did not occur during *No-Sonar*. Avoidance responses started during the ramp-up period (i.e. the first 5 minutes) in nearly all sessions with avoidance (8/9). For the one case where avoidance started during a full-power period (mn11_165e and f, session #3; Fig. A17), the received SPL was still increasing and source-whale range decreasing at the onset of response. The interaction between avoidance and presence of feeding behaviour prior to avoidance was not retained in the final model, indicating little support for an effect of feeding context on ramp-up efficacy in humpback whales.

Attraction to the source (Fig. 5.2C) occurred during 2 sessions with and 2 sessions without ramp-up (Fig. 5.4), and these attraction responses increased received levels on

average (+3 dB for SPL_{max}; +2 dB for SEL_{cum}) compared to sessions without attraction. The minimum source-whale distance (R_{min}) ranged from 61 to 998 m (Fig. 5.4C). R_{min} was increased by 145 m during *RampUp*, i.e. *RampUp1* and *RampUp2* combined, compared to *No-RampUp* (table 2). As expected, R_{min} was greater (+251 m) and smaller (−175 m) during sessions that included avoidance and attraction, respectively (Table 5.2).

DISCUSSION

Adaptive management of environmental impact requires evaluation of mitigation methods such as shut-down and ramp-up (Dolman et al., 2009). These results represent the first experimental examination of ramp-up used by several navies. The humpback whale was considered an appropriate model species for this purpose as they are relatively easy to find, tag and track, which improved the chances of obtaining a sufficient sample size. Humpbacks were generally not strongly responsive to 1.3-2 kHz sonar, and therefore ramp-up was not very effective for this species in general. In the 60% of *RampUp1* sessions where animals avoided the sound source, which is the assumption behind ramp-up (von Benda-Beckmann et al., 2014), received SPL_{max} and SEL_{cum} was significantly reduced by 6 dB and 2 dB, respectively. In a qualitative comparison between six species of cetacean, humpback whales were found to be relatively unresponsive to naval sonar (Sivle et al., 2015b), and thus it is expected that ramp-up should be more effective with other, more responsive species (von Benda-Beckmann et al., 2014).

Humpback whales have a global distribution and make long-distance migrations (Jackson et al., 2014; Stevick et al., 2011) during which they are likely to encounter anthropogenic disturbances (Rosenbaum et al., 2014). Their distribution during important behaviours such as feeding and mating is concentrated at inshore and continental shelf waters, which overlaps with that of naval sonar activity (Christensen et al., 1992). In the last two decades, several navies have started using lower frequency (≤2 kHz) active sonar systems for longer-range detections of submarines (Scott, 2015). Behaviour is highly context dependent, and contextual variables such as the novelty of a stimulus, along with the loudness and temporal aspects of the sound, can affect

responsiveness (Ellison et al., 2012). It is therefore possible that the observed behavioural responses of the whales were influenced by their history of experience with long-range sonar.

Another context variable that may influence the behavioural response (Richardson et al., 1995) and thus the effectiveness of ramp-up is the relative motion of the source vessel. The source vessel's line-of-approach was directly at the whale during the experiments, and animals may have responded differently if the transmitting source was stationary or moving away. It also meant that inaccuracies in the selected course and start time had relatively large effects on the received levels and were responsible for significant variation in the data set (Fig. 5.4). However, a sonar vessel that is approaching is more likely to put animals at risk of physiological damage, and the purpose of ramp-up is to reduce the risk for whales that happen to be directly in the path of the source vessel before it starts a full-level sonar operation (von Benda-Beckmann et al., 2014).

It should be emphasised that the main comparisons here were between the two *RampUp* sessions, and avoidance, attraction, or lack of responses observed in the sessions. The data for all but two *No-RampUp* sessions were based on modelling of sounds that would have been received if the source had been transmitting when the vessel was just passing the whale. During approaches with rapidly changing distances between the source and the whale, received SEL_{cum} primarily depends upon the highest single-pulse SPL received. The whale subject would not have had time to move enough to affect SEL_{cum} much if it reacted to these high-SPL pings. Thus, the use of full-level sonar sessions without ramp-up was not necessary, and arguably unethical, as modelling from *No-Sonar* sessions was adequate to estimate the SEL_{cum} that would have been received if the sonar had been transmitting. These types of experiments at sea are difficult and expensive to conduct, which means that compromises have to be made during analysis regarding sample size and combinations of conditions that can be tested. *No-Sonar* was always presented first and the two real *No-RampUp*s were presented third (Table 3.1), so potential order effects in stimulus presentation may have occurred.

Multiple parameters characterise the transmission scheme of a sonar ramp-up (e.g. total duration, pulse duration, inter-pulse interval, speed of source) and at present different ramp-up schemes are in operation. Only test one scheme could be tested experimentally; therefore, the ramp-up scheme was selected based upon considerations of operational relevance and a quantitative assessment of risk of effects on hearing impairment. The model framework described in (von Benda-Beckmann et al., 2014) was applied with parameter values adjusted for humpback whales and sonar source used (see Kvadsheim et al., 2011, for details). The pulse duration for ramp-up was also shorter than for full-power based on the premise that this reduced SEL_{cum} and thus risk of causing hearing damage but did not affect how loud the pulses were perceived by the whales (von Benda-Beckmann et al., 2014). Audibility and perceived loudness are known to increase up to signal durations of ~0.3 s in humans (Zwislocki, 1969), with comparable findings in other mammals tested to date (Kastelein et al., 2010; Richardson et al., 1995). Unfortunately, no such information is available for baleen whales. Even in the absence of a loudness effect the differential pulse duration might have had an effect on the responsiveness of the animals. However, pulse duration was just one of the parameters of the tested ramp-up, and extrapolations of the results of this chapter to other ramp-up schemes should always be done with great care.

The sonar source used in this study was previously tested as a prototype sonar system on operational Royal Netherlands Navy frigates. Its maximum SL of 214 dB re 1µPa m falls in the low end of the range of operational tactical naval sonars (210-240 dB re 1µPa m; Ainslie, 2010). The concept of ramp uses transmissions at reduced SL to decrease the exposure at the animal, and therefore the full-power SL was likely not a critical factor in determining whether or not ramp-up is effective. Theoretical modelling predicted that the absolute reduction in received level should be greater for higher maximum SLs because animals are more likely to respond at greater distances (von Benda-Beckmann et al., 2016a). The effectiveness of sonar ramp-up should also be determined by factors other than SL such as pulse interval, ship speed, animal swim speed, and whether the source is moving or stationary (Ainslie and von Benda-Beckmann, 2013; von Benda-Beckmann et al., 2014, 2016). Ramp-ups for other types

of sources likely have different considerations because animals may respond differently to these sounds (e.g. McCauley et al., 2000).

The sound source produced the relatively high received levels at the whale that were needed to invoke the behavioural responses; the highest received SEL_{cum} recorded by the DTAG was 181 dB re 1μPa^2 s (Fig. 5.1). This value lies substantially below a commonly-used criterion for temporary hearing loss onset (195 dB re 1μPa^2 s; Southall et al., 2007). Little is known about the hearing of baleen whales, and it cannot be excluded that some of the experimental subjects experienced small temporary reductions of hearing sensitivity after *RampUp1*, which could have reduced their responsiveness during *RampUp2*. However, a small shift in hearing sensitivity should have recovered after one hour. The lower responsiveness during *RampUp2* was therefore more likely caused by habituation, e.g. as a result of a change in the level of perceived risk associated with the sonar stimulus. Although an effect of the modest sample size cannot be ruled out (only 6 and 3 groups showed avoidance during *RampUp1* and *RampUp2*, respectively), parallel analyses of the same dataset found evidence for a general tendency of habituation also in other aspects of the humpbacks' behavioural responses (Sivle et al., 2015b). The humpback whales showed much stronger avoidance responses, and never attraction, when killer whale sounds were played at lower received levels than the sonar, indicating their ability to respond more strongly, and differences in perceived risk between the two stimuli (Curé et al., 2015; Sivle et al., 2015b).

The dataset used here included three potential mother-calf pairs, as these pairs were composed of an adult and a smaller-sized individual that remained closely associated with each other throughout the tracking record (Curé et al., 2015). One calf (Fig. 5.3C) was substantially smaller than the other two (Figs. 5.3A,B). The whale (mn12_180b) that was accompanied by this small calf responded during both *RampUp* sessions with an unusual 3-dimensional avoidance response, which included a descent to >100 m depth, without any signs of habituation (Figs. 5.2B, A24). Dive behaviour of the calf was not recorded as the calf was not tagged; however, this animal was always in close proximity with its presumed mother when she surfaced, and humpback whale calves are

likely to ‘follow’ the movements of their mothers under water (Tyson et al., 2012). Although anecdotal, this observation illustrates an important point: some subsets of a population, e.g. mother-calf pairs (McCauley et al., 2000) with very young calves and some taxa, e.g. beaked whales (DeRuiter et al., 2013; Miller et al., 2015a; Stimpert et al., 2014; Tyack et al., 2011), are more responsive than others, and ramp-up is most effective for these behaviourally sensitive animals (von Benda-Beckmann et al., 2014). This holds particularly true for sources that are moving rapidly relative to the animal, such as towed and hull-mounted naval sonars, as rapidly approaching ships give the animals less time to respond and move away (von Benda-Beckmann et al., 2016a). That the other two 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), but might also be caused by individual variation.

The observed heterogeneity in behavioural responsiveness to sonar suggests that both between- and within-species variation should be taken into account when performing environmental risk assessments and evaluation of risk mitigation measures. Similar heterogeneity was found in studies with other humpback whales (Dunlop et al., 2013; Maybaum, 1989; Miller et al., 2000) and blue whales (Goldbogen et al., 2013b). In contrast to the study of Goldbogen et al. (Goldbogen et al., 2013b), the behavioural state (feeding/non-feeding) of the whales was not an important predictor in this chapter. In 13% of sessions, the humpback whales appeared to be attracted to the research vessel or sonar stimulus, which increased the sound levels received by the whales. Positive reactions of humpback whales to whale-watching have been reported (Watkins, 1986). Although whale-watching does generally not occur in the study area, the whales might have experienced whale watching on their breeding grounds. In addition, juveniles and different cetacean species, e.g. northern bottlenose whales (*Hyperoodon ampullatus*), are sometimes drawn to boats and novel stimuli (Miller et al., 2015a). The current study demonstrates that such attraction behaviour reduces the effectiveness of ramp-up compared to responses of more skittish animals that are likely to respond quickly and move away.

Conclusions and recommendations

Ramping up the source level of a 1.3-2 kHz sonar to the maximum of 214 dB re 1μPa m was not an effective method to reduce received levels for humpback whales, because humpback whales in general were not very responsive to such signals (Sivle et al., 2015b). However, clear within-population variation in behavioural responsiveness occurred within the sample; some whales avoided the source while others did not, or were attracted to it. Based upon an analysis that separated these different types of responses, more specific conclusions were drawn about when and how ramp-up is effective as a mitigation measure. I conclude that ramp-up can reduce received levels and thereby the risk of physiological damage in particularly responsive animals such as subsets of the humpback whales population as well as other species. The observation of strong avoidance without habituation in the presumed cow with a small calf strengthens this conclusion. I therefore encourage the use of ramp-up of naval sonar in breeding grounds and other areas with such behaviourally sensitive animals. However, when animals have strong motivations not to move away from their current location, ramp-up cannot be effective. Other operational mitigation procedures, e.g. warning sounds prior to sonar pulses (Nachtigall and Supin, 2013), should be considered in addition to ramp-up. If the sound following a warning is aversive, then animals may become conditioned to avoid it or to reduce their hearing sensitivity. Further, some operational sonars may have equipment limitations that preclude the use of ramp-up (Dolman et al., 2009); therefore, I recommend that new sonars are designed to include more flexibility in transmission schemes and signal types.

REFERENCES

Ainslie, M. A. (2010). Principles of Sonar Performance Modeling. Chichester, UK: Springer-Praxis.

Ainslie, M. A. and von Benda-Beckmann, A. M. (2013). Optimal soft start and shutdown procedures or stationary or moving sound sources. Proc. Meet. Acoust. 17, 070077.

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

Barlow, J. and Gisiner, R. (2006). Mitigating, monitoring and assessing the effects of anthropogenic sound on beaked whales. J. Cetacean Res. Manag. 7, 239–249.

Barton, K. (2015). Package "MuMIn": Multi-model inference.

Boyd, I., Brownell, B., Cato, D., Clark, C., Costa, D., Evans, P., Gedamke, J., Gentry, R., Gisiner, R., Gordon, J., et al. (2008). The effects of anthropogenic sound on marine mammals. A draft research strategy. (ed. Connolly, N.) Marine Board-European Science Foundation.

Brandt, M. J., Diederichs, A., Betke, K. and Nehls, G. (2011). Responses of harbour porpoises to pile driving at the Horns Rev II offshore wind farm in the Danish North Sea. Mar. Ecol. Prog. Ser. 421, 205–216.

British Oceanographic Data Centre (2010). The GEBCO_08 Grid.

Christensen, I., Haug, T. and Øien, N. (1992). Seasonal distribution, exploritation and present abundance of stocks of large baleen whales (Mysticeti) and sperm whales (*Physeter macrocephalus*) in Norwegian and adjecent waters. ICES J. Mar. Sci. 49, 341–355.

Compton, R., Goodwin, L., Handy, R. and Abbott, V. (2008). A critical examination of worldwide guidelines for minimising the disturbance to marine mammals during seismic surveys. Mar. Pol. 32, 255–262.

Curé, C., Doksaeter 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.

DeRuiter, S. L., Southall, B. L., Calambokidis, J., Zimmer, W. M. X., Sadykova, D., Falcone, E. A., Friedlaender, A. S., Joseph, J. E., Moretti, D., Schorr, G. S., et al. (2013). First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. Biol. Lett. 9, 20130223.

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.

dos Santos, M. E., Couchinho, M. N., Rita Luis, A. and Goncalves, E. J. (2010). Monitoring underwater explosions in the habitat of resident bottlenose dolphins. J. Acoust. Soc. Am. 128, 3805–3808.

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.

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.

Frankel, A. S. and Clark, C. W. (2000). Behavioral responses of humpback whales (*Megaptera novaeangliae*) to full-scale ATOC signals. J. Acoust. Soc. Am. 108, 1930–1937.

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.

Hardin, J. W. and Hilbe, J. M. (2003). Generalized Estimating Equations. Boca Raton, FL, USA: Chapman and Hall/CRC Press.

Harris, C. M., Sadykova, D., DeRuiter, S. L., Tyack, P. L., Miller, P. J. O., Kvadsheim, P. H., Lam, F. P. A. and Thomas, L. (2015). Dose response severity functions for acoustic disturbance in cetaceans using recurrent event survival analysis. Ecosphere 6, art236.

Højsgaard, S., Halekoh, U. and Yan, J. (2014). Package "geepack": Generalized estimating equation package.

Houser, D. S., Martin, S. W. and Finneran, J. J. (2013a). Exposure amplitude and repetition affect bottlenose dolphin behavioral responses to simulated mid-frequency sonar signals. J. Exp. Mar. Bio. Ecol. 443, 123–133.

Houser, D. S., Martin, S. W. and Finneran, J. J. (2013b). Behavioral responses of California sea lions to mid-frequency (3250-3450 Hz) sonar signals. Mar. Environ. Res. 92, 268–278.

Jackson, J. A., Steel, D. J., Beerli, P., Congdon, B. C., Olavarria, C., Leslie, M. S., Pomilla, C., Rosenbaum, H. and Baker, C. S. (2014). Global diversity and oceanic divergence of humpback whales (*Megaptera novaeangliae*). Proc. R. Soc. B 281, 20133222.

Jefferson, T. A., Hung, S. K. and Wursig, B. (2009). Protecting small cetaceans from coastal development: Impact assessment and mitigation experience in Hong Kong. Mar. Pol. 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.

Kastelein, R. A., Hoek, L., de Jong, C. A. and Wensveen, P. J. (2010). The effect of signal duration on the underwater detection thresholds of a harbor porpoise (*Phocoena phocoena*) for single frequency-modulated tonal signals between 0.25 and 160 kHz. J. Acoust. Soc. Am. 128, 3211–3222.

Kastelein, R. A., Steen, N., de Jong, C., Wensveen, P. J. and Verboom, W. C. (2011). Effect of broadband-noise masking on the behavioral response of a harbor porpoise (*Phocoena phocoena*) to 1-s duration 6-7 kHz sonar up-sweeps. J. Acoust. Soc. Am. 129, 2307–2315.

Kastelein, R. A., Steen, N., Gransier, R., Wensveen, P. J. and de Jong, C. A. (2012). Threshold received sound pressure levels of single 1-2 kHz and 6-7 kHz up-

sweeps and down-sweeps causing startle responses in a harbor porpoise (*Phocoena phocoena*). J. Acoust. Soc. Am. 131, 2325–2333.

Kvadsheim, P., Lam, F.-P., Miller, P., Doksaeter, 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 Norwegain waters - 3S-2011 cruise report. FFI-rapport 2011/01289.

Kvadsheim, P., Lam, F.-P., Miller, P., Wensveen, P., Visser, F., Sivle, L., Kleivane, L., Curé, C., Ensor, P., van Ijsselmuide, S., et al. (2012). Behavioural responses of cetaceans to naval sonar signals in Norwegian waters – the 3S-2012 cruise report. FFI-rapport 2012/02058.

Kvadsheim, P. H., Miller, P. J. O., Tyack, P. L., Sivle, L. D., Lam, F. P. A. and Fahlman, A (2012). Estimated tissue and blood N(2) levels and risk of decompression sickness in deep-, intermediate-, and shallow-diving toothed whales during exposure to naval sonar. Front. Physiol. 3, 125.

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

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.

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

Miller, P. J. O., Kvadsheim, P. H., Lam, F. P. A., Tyack, P. L., Curé, C., DeRuiter, S. L., Kleivane, L., Sivle, L. D., van IJsselmuide, S. P., Visser, F., et al. (2015). First indications that northern bottlenose whales are sensitive to behavioural disturbance from anthropogenic noise. R. Soc. Open Sci. 2, 140484.

Mooney, T. A., Nachtigall, P. E. and Vlachos, S. (2009). Sonar-induced temporary hearing loss in dolphins. Biol. Lett. 5, 565–567.

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.

Nowacek, D. P., Bröker, K., Donovan, G., Gailey, G., Racca, R., Reeves, R. R., Vedenev, A. I., Weller, D. W. and Southall, B. L. (2013). Responsible practices for minimizing and monitoring environmental impacts of marine seismic surveys with an emphasis on marine mammals. Aquat. Mamm. 39, 356–377.

Porter, M. B. and Bucker, H. P. (1987). Gaussian beam tracing for computing ocean acoustic fields. J. Acoust. Soc. Am. 82, 1349–1359.

Richardson, W. J., Greene, C. R., Malme, C. I. and Thomson, D. H. (1995). Marine Mammals and Noise. San Diego, CA: Academic Press.

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.

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.

Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene Jr., C. R., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., et al. (2007).

Marine mammal noise exposure criteria: Initial scientific recommendations. Aquat. Mamm. 33, 411–521.

Stevick, P. T., Neves, M. C., Johansen, F., Engel, M. H., Allen, J., Marcondes, M. C. C. and Carlson, C. (2011). A quarter of a world away: female humpback whale moves 10 000 km between breeding areas. Biol. Lett. 7, 299–302.

Stimpert, A. K., DeRuiter, S. L., Southall, B. L., Moretti, D. J., Falcone, E. A., Goldbogen, J. A., Friedlaender, A., Schorr, G. S. and Calambokidis, J. (2014). Acoustic and foraging behavior of a Baird's beaked whale, *Berardius bairdii*, exposed to simulated sonar. Sci. Rep. 4, 7031.

Stone, C. J. (2015). Marine mammal observations during seismic surveys from 1994-2010. JNCC Report no. 463a.

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.

Tasker, M. L., Amundin, M., Andre, M., Hawkins, A., Lang, W., Merck, T., Scholik-Schlomer, A., Teilmann, J., Thomsen, F., Werner, S., et al. (2010). Marine strategy framework directive: Task group 11 report: Underwater noise and other forms of energy. Luxembourg: European Union and International Council for the Exploration of the Sea.

Tyack, P. (2009). Acoustic playback experiments to study behavioral responses of free-ranging marine animals to anthropogenic sound. Mar. Ecol. Prog. Ser. 395, 187–200.

Tyack, P., Gordon, J. and Thompson, D. (2003). Controlled-exposure experiments to determine the effects of noise on marine mammals. Mar. Technol. Soc. J. 37, 39–51.

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

Tyson, R. B., Friedlaender, A. S., Ware, C., Stimpert, A. K. and Nowacek, D. P. (2012). Synchronous mother and calf foraging behaviour in humpback whales *Megaptera novaeangliae*: insights from multi-sensor suction cup tags. Mar. Ecol. Prog. Ser. 457, 209–220.

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, Advances in Experimental Medicine and Biology, pp. 1197–1203.

Watkins, W. A. (1986). Whale reactions to human activities in Cape Cod waters. Mar. Mamm. Sc. 2, 251–262.

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. (2015). 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.

Zwislocki, J. J. (1969). Temporal summation of loudness: an analysis. J. Acoust. Soc. Am. 46, 431–441.

TABLES & FIGURES

Table 5.1. Details of the stepwise variable selection procedure conducted on the extended GEE models for the session's received maximum sound pressure level (SPL_{max}), received cumulative sound exposure level (SEL_{cum}), and minimum source-whale range (R_{min}). The test statistic (χ^2) and p-value of the factor covariate that was dropped at a given step are highlighted in bold typescript. The final, most-parsimonious model was achieved for each response variable after 5 steps.

Response variable	**Predictor variable**	**Step 1**		**Step 2**		**Step 3**		**Step 4**		**Step 5**	
		χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
SPL_{max}	SessionType	1.4	0.239	1.4	0.239	**1.4**	**0.239**				
	SessionOrder	**1.1**	**0.286**								
	Attraction	4.3	0.037	7.8	0.005	7.8	0.005	8.5	0.004	8.5	0.004
	Avoidance	11.6	0.001	8.8	0.003	8.8	0.003	9.1	0.003	9.1	0.003
	AvoidOrder	4.0	0.046	6.5	0.000	6.5	0.011	6.3	0.012	6.3	0.012
	Feeding	2.4	0.124	2.4	0.011	2.4	0.125	**2.5**	**0.115**		
	Avoidance: Feeding	1.2	0.281	**1.0**	**0.329**						
SEL_{cum}	SessionType	**0.0**	**0.948**								
	SessionOrder	2.7	0.100	2.1	0.145	**2.1**	**0.145**				
	Attraction	12.7	0.000	13.0	0.000	13.0	0.000	9.1	0.003	9.1	0.003
	Avoidance	8.1	0.004	7.9	0.005	7.9	0.005	6.4	0.012	6.4	0.012
	AvoidOrder	5.4	0.020	2.5	0.116	2.5	0.116	9.8	0.002	9.8	0.002
	Feeding	2.4	0.122	2.7	0.101	2.7	0.101	**2.3**	**0.134**		
	Avoidance: Feeding	0.6	0.453	**0.7**	**0.420**						
R_{min}	SessionType	5.3	0.021	5.3	0.021	5.3	0.021	5.3	0.021	5.3	0.021
	SessionOrder	1.2	0.280	**1.2**	**0.280**						
	Attraction	5.8	0.016	5.8	0.016	7.0	0.008	7.0	0.008	7.0	0.008
	Avoidance	9.5	0.002	9.5	0.002	8.1	0.004	8.1	0.004	8.1	0.004
	AvoidOrder	1.6	0.244	1.6	0.200	2.3	0.133	**2.3**	**0.133**		
	Feeding	3.0	0.084	3.0	0.084	**1.9**	**0.166**				
	Avoidance: Feeding	**0.0**	**0.873**								

Table 5.2. Statistics for the basic GEE models with covariates SessionType and SessionOrder, and for the final (most-parsimonious) GEE models. Only significant factor covariates, highlighted in bold, were interpreted.

model	**covariate (reference level)**	**received maximum SPL**				**received cumulative SEL**			
		estimate	*SE*	*Z*	*p*	*estimate*	*SE*	*Z*	*p*
basic	intercept	168.2	1.0			175.3	0.8		
	SessionType (*No-RampUp*)	−2.2	1.2	3.7	0.055	−1.1	0.6	3.3	0.069
	SessionOrder (*RampUp1*)	2.1	1.9	1.2	0.284	2.2	1.2	3.3	0.071
final	intercept	168.3	0.7			175.6	0.5		
	Attraction (absence)	**3.5**	**1.5**	**5.1**	**0.024**	**2.3**	**0.5**	**18.4**	**<0.0001**
	Avoidance (absence)	**−5.7**	**1.6**	**12.2**	**<0.001**	**−3.5**	**1.0**	**12.2**	**<0.001**
	AvoidanceOrder (during *RampUp1*)	**3.6**	**1.4**	**6.6**	**0.010**	**3.9**	**1.1**	**12.7**	**<0.001**
	SessionType (*No-RampUp*)								

Table 5.2. Continued.

model	**covariate (reference level)**	**minimum source-whale range**			
		estimate	*SE*	*Z*	*p*
basic	intercept	247.1	33.9		
	SessionType (*No-RampUp*)	**144.7**	**60.1**	**5.8**	**0.016**
	SessionOrder (*RampUp1*)	−127.1	107.9	1.4	0.239
final	intercept	287.4	39.8		
	Attraction (absence)	**−174.9**	**87.3**	**4.0**	**0.045**
	Avoidance (absence)	**251.0**	**78.4**	**10.2**	**0.001**
	AvoidanceOrder (during *RampUp1*)				
	SessionType (*No-RampUp*)	−80.7	51.9	2.4	0.120

Figure 5.1. Received sound exposure levels (SELs) of all the transmitted 1.3-2 kHz sonar pulses. Measured levels were calculated from the acoustic recordings made by the DTAGs (green) and modelled levels were based on propagation losses (PLs) predicted using BELLHOP (black). Single-pulse SELs are shown as dots and cumulative SEL for the whole exposure session is shown as a line. The value in the bottom right corner of each panel represents the difference in cumulative SEL at the end of the session (i.e. measured level – predicted level).

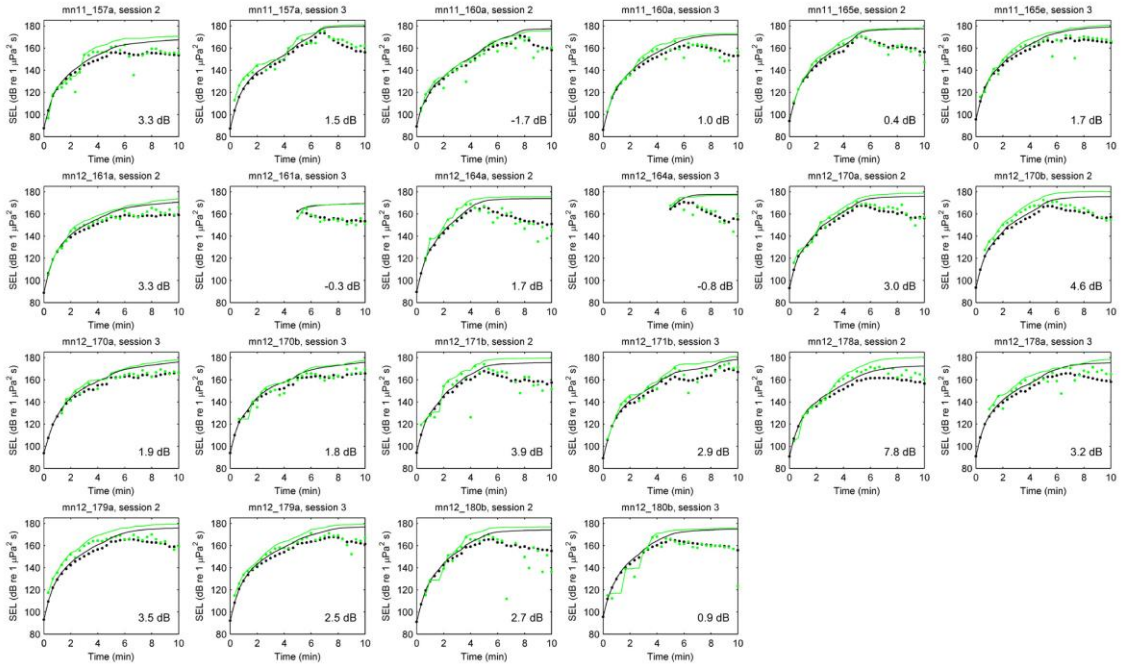


Figure 5.2. Three typical examples of humpback whale movement before, during, and after sonar exposures: (A) no response (*No-Sonar*; mn11_158a), (B) avoidance (*RampUp2*; mn12_180b), and (C) attraction and then avoidance (*RampUp1*; mn12_164a). The position of the source (dot) and position of the research vessel towing the source (circle) are indicated for each pulse transmission. Note the difference in depth scales. See Appendix II for plots of all sessions.

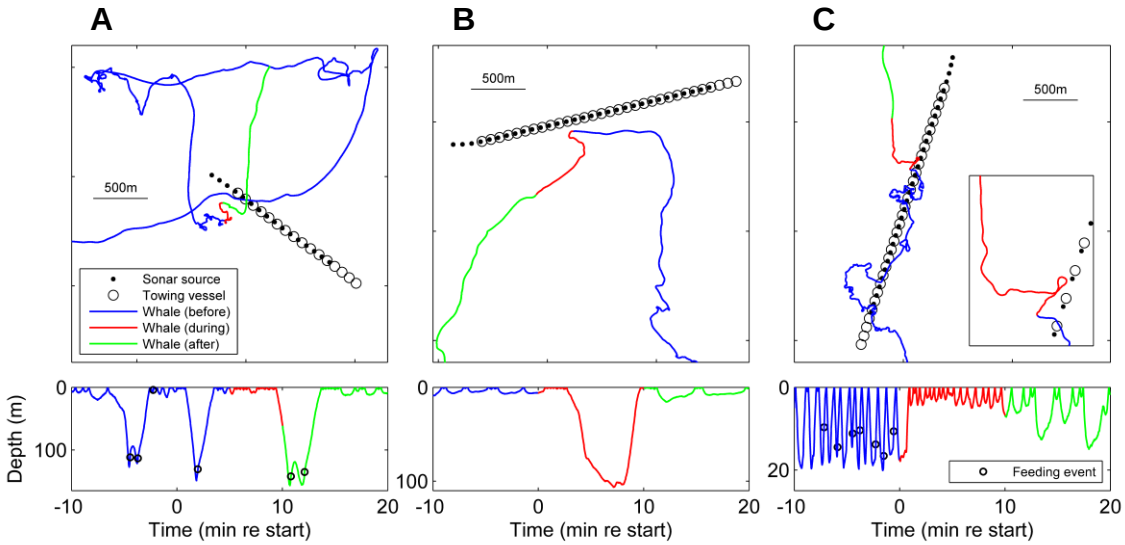


Figure 5.3. Potential mother-calf pairs: (A) focal whale mn11_160a and an untagged adult, (B) focal whale mn11_165e and calf mn11_165f, and (C) focal whale mn12_180b and a small untagged calf.

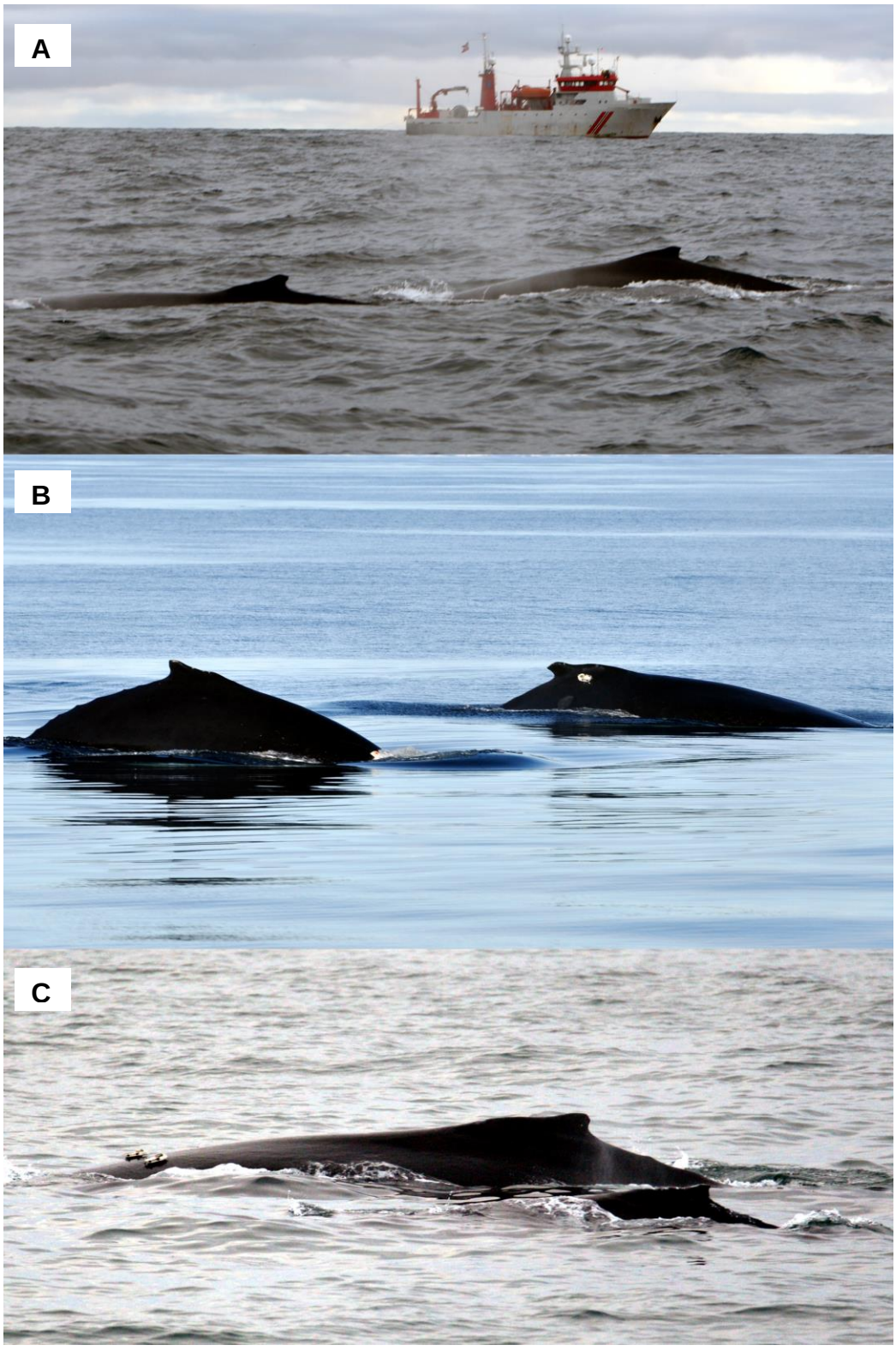
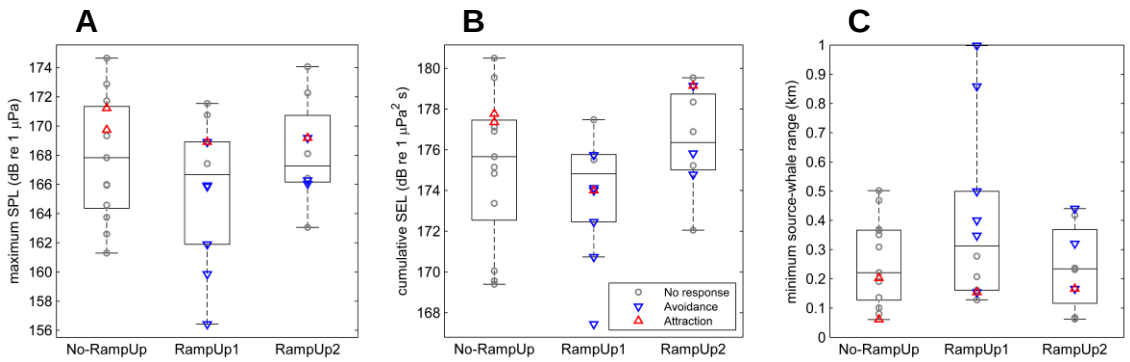


Figure 5.4. Sound levels received by the tagged whales, expressed as (A) SPL_{max} and (B) SEL_{cum}, and the (C) minimum source distance (R_{min}) during *No-RampUp*, and first *RampUp* and second *RampUp* sessions. Boxplots are overlaid by raw data markers indicating the presence of avoidance or attraction.



Chapter 6

General discussion

The overall goals of this thesis were 1) to collect scientific evidence that can inform criteria for allowable levels of naval sonar and 2) to test the effectiveness of an operational method that is currently in use for mitigating effects of naval sonar on marine mammals. Management of risk of anthropogenic noise effects is severely hindered by the lack of quantitative information on the hearing abilities of marine animals and their behaviour in natural conditions and during anthropogenic disturbance.

Knowledge about behavioural and physiological responses of marine mammals to high-intensity naval sonar has significantly contributed to the scientific debate about the impacts of anthropogenic noise. In this thesis, I have sought to explore specifically the potential effects of naval sonar by means of novel experimental and analytical methods and using both wild and captive cetaceans. The development of novel methods is essential in cetacean research because their fully aquatic lifestyle makes it challenging to study these animals in the wild. Research on cetaceans is hampered further by the fact that only a handful of animals of a limited number of species are available for captive research.

SYNTHESIS OF RESULTS

The development of the track reconstruction method (Chapter 2) was motivated by the need for accurate position estimates during the experimental tests of the effectiveness of ramp-up (Chapter 5). Short distances between the sonar source and the tagged humpback whales during the controlled exposure experiments were expected *a priori*, and therefore the distance estimates and their variability were likely to significantly affect the modelled received sound levels and the conclusions drawn from these experiments. But the track reconstruction method was also instrumental for the identification of avoidance and attraction responses used in the analyses in Chapters 3 and 5. Two horizontal avoidance responses may have been missed without the use of

the fine-scale movement tracks, which likely had an effect on the dose-response function for this type of response (Figs. 3.9 and 3.10).

Chapter 3 quantified the probability of avoidance onset and feeding cessation as a function of received sonar dose for 18 exposure sessions with 10 humpback whales. The probability of behavioural response may depend upon the frequency of the sound stimulus, as all animals perceive sound differently depending on frequency. Extrapolating the dose-response functions presented here for 1.3-2 kHz upsweep sonar signals towards other signal frequencies is therefore not straight-forward, and the mechanism by which response probability should be distributed across the hearing range is not known. For example, response probabilities could be extrapolated following the assumption that only the received SPL (Fig. 6.1A) or the sensation level (Fig. 6.1B) drives the responses. The response probability at each signal frequency could also be assumed to increase from 0 at the absolute hearing threshold to 1 at a fixed upper limit SPL where the sound becomes unbearably loud (Fig. 6.1C). At least conceptually, this latter approach is probably the most consistent with current knowledge on loudness perception in normal-hearing humans and other mammals, including with the newly-derived equal latency contours for a harbour porpoise (Chapter 4).

The effectiveness of ramp-up for moving sonar vessels depends greatly upon the behavioural responsiveness of the exposed animal. The sonar dose at the onset of response (i.e. the dose-response function), but also the swimming speed and direction, and diving behaviour will determine if an avoidance is effective enough for ramp-up to reduce risks in animals located near the sound source. The ramp-up of sonar source levels did not result in an overall decrease in received sound level or increase in minimum source-whale distance in the experimental tests with humpback whales (Chapter 5), which is in line with the relatively high avoidance thresholds in these animals (Fig. 3.10) compared to those of several other cetacean species tested (Fig. 3.12). This suggests that the dose-response functions reported for avoidance behaviours of other species can be used to provide indications of how effective ramp-up will be for these specific species.

The SPL dose-response function for avoidance by humpback whales had a p_{50} value of 166 dB re 1 μPa (Fig. 3.10). Von Benda-Beckmann et al. (2014) used a theoretical modelling approach to investigate the effectiveness of sonar ramp-up in terms of reducing the area in which animals are experiencing sound doses high enough to cause temporary threshold shift (TTS) or permanent hearing loss. These authors used data on natural behaviour of killer whales and their responses to sonar to inform the behaviour of the simulated animals, but many of their results can be generalised to other marine mammal species. Von Benda-Beckmann et al. (2014) also investigated the effects of several dose-response functions on the efficacy of the ramp-ups; including one with a very similar p_{50} value and comparable steepness as the humpback whale SPL function (see Fig. 3 of that article). Their simulation study showed that ramp-up prior to a full-power sonar operation was generally not effective if the avoidance thresholds of the exposed population of simulated animals followed this particular dose-response function.

In contrast, ramp-up of sonar was effective if the simulated animals were overall more responsive (von Benda-Beckmann et al., 2014). Behavioural responses can be influenced by many factors relating to the environmental or individual context of the exposed animal (Ellison et al., 2012; Radford et al., 2016; Richardson et al., 1995). Chapters 3 and 5 in this thesis provided evidence that the novelty of the sound stimulus, the animal being in a non-feeding state, and the presence of a small calf increased responsiveness for humpbacks (although the latter conclusion is based on only one mother-calf pair). Therefore, the effectiveness of ramp-up may be higher for a particular subset of the humpback whale population, and similar heterogeneity in responsiveness can likely be found in other species and populations.

Dose-response functions specific to behavioural state (feeding / non-feeding) and short-term exposure history were presented (Fig. 3.10). Such context-specific dose-response functions may improve predictions of behaviour responses if they are combined with information about temporal and spatial variability in the prevalence of these certain contexts. For example, humpback whales around the western Antarctic Peninsula predominantly feed at night late in the feeding season (Friedlaender et al., 2013), while

humpback whale mother-calf pairs may use nearshore waters such as Exmouth Gulf, Australia as resting areas (McCauley et al., 2000). Responsiveness to novel stimuli and changes in tolerance levels over time may be especially important for understanding the potential differences between areas with much naval activity and areas that are relatively pristine to sonar exposure (Miller et al., 2015b). However, this part of the analysis also highlighted an important limitation of tagging studies with cetaceans, which generally have small sample sizes. Clear differences amongst the context-specific dose-response functions were apparent (Fig. 3.9), but the samples sizes of the subsets were too small for the derivation of robust conclusions about the effects of the contexts.

METHODOLOGICAL CONSIDERATIONS

Very few cetacean studies have used Fastloc-GPS in combination with other movement sensors (but see McKenna et al., 2015). This will likely change when more multi-sensor tags will incorporate Fastloc-GPS into their sensor package. As a result, the demand for integrative methods for multi-variate movement data is likely to increase in the near future. Fine-scale movement tracks of cetaceans have the potential to fill interesting knowledge gaps, e.g. about predator movements in relation to prey fields, dynamics of group movement, impacts of human disturbance, and foraging effort and success.

The track reconstruction method developed in Chapter 2 had a satisfactory performance for all humpback whale data sets, even though they included different degrees of data coverage, data quality and apparent environmental conditions. Good overall performance is vital for such analytical methods because the data collected by inertial, speed and positioning sensors may be affected by various sources of error. The model framework was written in BUGS language and models were fitted in the standard software program JAGS using Bayesian methods. As a result, model fitting may become relatively slow when tracks contain many Fastloc-GPS positions (Table 2.1). Other disadvantages are the need for pre- and post-processing of the time series, and some tendency to underestimate the uncertainty in location at times between position fixes. However, adapting the model framework is relatively easy, and the track reconstruction method has potential application beyond humpback whales and beyond the data recording systems used.

The noise separation method of von Benda-Beckmann et al. (2016b) was used to assess if noise levels recorded on the DTAG were affected by flow noise (Chapter 3). This method proved to be a useful tool, as it enabled the estimation of dose-response functions for signal-to-noise ratio. The method was not yet available when the track reconstruction model was developed; it could have been used to obtain more reliable estimates of speed-through-water when the tagged whale was very close to the surface. Estimates of the ambient noise are often a missing component in behavioural response studies with tagged animals, so the method may be used in future studies to more fully describe the ambient noise field.

The use of a simple optical sensor system to measure the underwater reaction times of the harbour porpoise was another novel technique used in this thesis (Fig. 4.3). The intensity of the infrared light was amplitude-modulated by an integrated circuit, making the detector impervious to disturbing external light. Reaction times recorded during a trial period were found to be affected by algal growth, so from then on the sensor was cleaned daily during the data collection period. The sensor system did not have components that may have suffered from changes in water temperature or mechanical fatigue and was found to recorded precise measurements of underwater reaction time to the sound stimuli, so this technique may be useful for future investigations.

IMPLICATIONS FOR MANAGEMENT

Quantitative data on the potential effects of anthropogenic sound on cetaceans remains limited (Nowacek et al., 2007), so the information reported here about dose-response functions for behaviour, loudness perception based on reaction times, and the effectiveness of sonar ramp-up may have important management implications. The reported dose-response functions and weighting functions can be used to inform noise exposure guidelines for behavioural impacts of tonal sounds on low- and high-frequency cetaceans. The results of the ramp-up study are especially relevant for establishing mitigation strategies for fast-moving sound sources. The detailed information on the movement responses of the humpback whales may also be useful; for example, for improving modelling efforts that are aimed at predicting the risk of

ship strikes (e.g. Bezamat et al., 2014) and predicting the number of marine animals that will be negatively affected by anthropogenic noise exposure (e.g. Frankel et al., 2016; Wartzok et al., 2012).

The similarity in dose-response functions between onset of avoidance and cessation of feeding suggested that horizontal displacement and reductions in feeding occurred at similar disturbance levels in humpback whales, even if these responses do not occur at the same time. This prediction is consistent with the severity scale of Southall et al. (2007), which assigned the same level of importance to these two response types. Comparison with published dose-response functions for other marine mammals indicated that the humpback whales were relatively unresponsive in general, so behavioural disturbance may be expected to occur at smaller distances and at higher received sound levels in this species compared to other marine mammals. A minority of humpback whales were predicted to respond at relatively low received levels; for example, about 20% of responses corresponded to received SPLs of <140 dB re 1 μPa. However, whether biologically significant impacts can be inferred from the short-term responses described here remains to be determined.

Compared to the first sonar exposure session, response thresholds measured during the second sonar exposure sessions were higher for most humpback whales but lower for the mother with the small calf. This difference suggested that the transmitting sonar source (source level: 214 dB re 1 μPa m) was generally perceived by humpback whales as having a low risk, but that behavioural responsiveness in some animals including mothers with young calves may escalate over time when the sonar exposure continues. Mothers with young calves may be particularly vulnerable to disturbance, so reducing impacts to these animals should be high priority for management of sonar effects and other conservation efforts.

A set of frequency weighting functions for marine mammals, known as “M-weighting”, have been widely used in the assessment of anthropogenic noise impacts (Southall et al., 2007). These weighting functions are based upon the functional hearing ranges of species groups and are relatively wide; they effectively apply no weighting over a wide

range of frequencies in which animals are expected to have some hearing sensitivity. As a result, M-weighting is conservative, i.e. over-protective towards the animals, when a given sound exposure is compared to an established risk criterion or safety limit, but it is not when applied to research data to obtain the risk criteria themselves (Tougaard et al., 2015b). The opposite may occur when audiogram-based frequency weighting is used, as this method is more likely to produce filter characteristics that are too narrow compared to the optimal weighting functions.

Tougaard et al. (2015) proposed that two different weighting functions could be applied, one for establishing the noise criterion from research and one for comparing the criterion to a proposed sound exposure. Ideally, sound exposure guidelines for marine mammals should include a single weighting function per species group, which matches the frequency response of the auditory system of the average animal. The natural choice would be a weighting function based upon equal loudness, analogous to A-weighting used for humans, but equal loudness contours currently only exist for one marine mammal (Finneran and Schlundt, 2011) and the collection of such data is highly impractical.

The general pattern of flattening of equal-loudness contours towards higher levels, seen in humans and *Tursiops*, matched the pattern in the equal latency contours of the harbour porpoise at frequencies below 40 kHz (Chapter 4), suggesting good correspondence between the perceived loudness of the sound stimulus and the reaction time it elicited in the animal for this frequency range. The derived weighting functions generally had a smaller low-frequency slope than the audiogram. Therefore, these equal latency data supported their use as proxies for equal loudness and the integration of these data in sound exposure criteria for marine mammals (as in the recent draft guidelines of NOAA, 2015). The comparisons of the equal latency contours with a small number of TTS onset thresholds and several behavioural response onset thresholds also tentatively supported the theoretical expectation that the equal loudness contours of the porpoise flatten towards higher perceived loudness levels.

In a recent study, Mulsow et al. (2015) derived the underwater equal latency contours of two bottlenose dolphins and in-air equal latency contours of three California sea lions. Increased compression of the equal latency contours was observed toward lower test frequencies in the sea lions, but not in the bottlenose dolphins. The reaction times generally tended to an asymptote with increasing SPL relatively quickly in both species, so the authors calculated the equal latency contours only at sensation levels up to ~40-50 dB. Similar patterns in reaction times have been occasionally reported for humans (e.g. Epstein and Florentine, 2006; Wagner et al., 2004), although most studies with humans and non-human animals have observed less pronounced asymptotes of reaction time (e.g. Marshall and Brandt, 1980; Pfingst et al., 1975; Stebbins, 1966; Wagner et al., 2004) (Fig. 4.5).

This difference may be caused by methodical differences including the species, stimulus, response types or psychoacoustic methods. For example, compared to the study on the harbour porpoise in Chapter 4, Mulsow et al. (2015) used shorter random waiting times and stimulus rise times, and generally narrower ranges of SPLs. In addition, the reaction times measured in the bottlenose dolphins were for vocal responses and not for motor responses, and these animals were significantly older (20-30 years) compared to the harbour porpoise (6-7 years) and the sea lions (2-5 years). Reaction times may be more easily affected by factors other than the loudness of the sound. Future studies on the influence of 'context variables', such as motivational state, physiological state, age, signal characteristics and background noise transients, on reaction times would greatly improve our understanding of the benefits and limitations of using frequency weighting based upon equal latency data.

Navies use ramp-up for moving source vessels to reduce risks of hearing injury in animals that are directly in the path of the vessel. Ramp-up therefore cannot be effective for animals at the side or behind the ship, and ramp-up is unlikely to reduce other types of impact. The most precautionary approach to naval sonar mitigation would probably comprehend a mitigation strategy that includes ramp-up and other protective measures that are implemented in the planning stage of a sonar exercise, in real-time during the operation, or after the sonar exercise is completed (Dolman et al., 2011). Besides ramp-

up, real-time operational methods may include reductions of sound levels at the source, warning sounds that are aimed at conditioning animals to avoid the source or reduce their hearing sensitivity during loud exposures (Nachtigall and Supin, 2013), and 'pre-watch periods' and 'shut-downs' based upon passive acoustic and/or visual monitoring of animals for the purpose of maintaining an 'exclusion zone' around the source. Monitoring using static acoustic recorders and from independent platforms such as other ships, gliders and autonomous underwater vehicles can be used to monitor animal presence over larger areas.

Evaluations of the effectiveness of risk mitigation methods for anthropogenic noise exposure to marine mammals are rarely conducted. Such evaluations are important for effectively balancing the costs and benefits of mitigation methods and to optimise mitigation strategies for a given noise-producing activity. Several variants of a mitigation method are generally in use. Therefore, the best approach for establishing the effectiveness of a mitigation method is likely to be a combination of carefully-designed experimental studies (Chapter 5; Götz and Janik, 2015), longitudinal (observational) studies (e.g. Stone, 2015) and computer simulations based on empirical data (Frankel et al., 2016; Leaper et al., 2015; von Benda-Beckmann et al., 2014).

Conclusions about the parameters that define an effective ramp-up for moving naval sonar vessels can be drawn from the experimental work in Chapter 5 and the theoretical work of von Benda-Beckmann et al. (2014, 2016a). In short, a ramp-up that effectively mitigates risks of physiological damage will be transmitted at a low ship speed so that the relative displacement of the animal is large. In addition, this ramp-up will have a relative short total duration, a low source level at the start of the ramp-up, relative short pulse durations to reduce the cumulative sound exposure level (SEL_{cum}) while maintaining the loudness of the single pulse, an aversive signal type such as an upsweep (Briefer, 2012; Kastelein et al., 2014), and an intermediate pulse interval that balances the contribution to the SEL_{cum} with the need to avoid long silences between successive pulses. Effective ramp-ups for stationary naval sources such as helicopter dipping sonars will have different considerations, and will generally include a longer total duration (Ainslie and von Benda-Beckmann, 2013).

REFERENCES

Ainslie, M. A. and von Benda-Beckmann, A. M. (2013). Optimal soft start and shutdown procedures or stationary or moving sound sources. Proc. Meet. Acoust. 17, 070077.

Bezamat, C., Wedekin, L. L. and Simões-Lopes, P. C. (2014). Potential ship strikes and density of humpback whales in the Abrolhos Bank breeding ground, Brazil. Aquat. Conserv. Mar. Freshw. Ecosyst. 1–14.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. J. Zool. 288, 1–20.

Dolman, S. J., Evans, P. G., Notarbartolo-di-Sciara, G. and Frisch, H. (2011). Active sonar, beaked whales and European regional policy. Mar. Pollut. Bull. 63, 27–34.

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.

Epstein, M. and Florentine, M. (2006). Reaction time to 1- and 4-kHz tones as a function of sensation level in listeners with normal hearing. Ear Hear. 27, 424–429.

Finneran, J. J. and Schlundt, C. E. (2011). Subjective loudness level measurements and equal loudness contours in a bottlenose dolphin (*Tursiops truncatus*). J. Acoust. Soc. Am. 130, 3124–3136.

Frankel, A. S., Ellison, W. T., Vigness-Raposa, K. J., Giard, J. L. and Southall, B. L. (2016). Stochastic modeling of behavioral response to anthropogenic sounds. In The Effects of Noise on Aquatic Life II, Advances in Experimental Medicine and Biology, pp. 321–329.

Friedlaender, A., Tyson, R., Stimpert, A., Read, A. and Nowacek, D. (2013). Extreme diel variation in the feeding behavior of humpback whales along the western Antarctic Peninsula during autumn. Mar. Ecol. Prog. Ser. 494, 281–289.

Götz, T. and Janik, V. M. (2015). Target-specific acoustic predator deterrence in the marine environment. Anim. Conserv. 18, 102–111.

Kastelein, R. A., Schop, J., Gransier, R., Steen, N. and Jennings, N. (2014). Effect of series of 1 to 2 kHz and 6 to 7 kHz up-sweeps and down-sweeps on the behavior of a harbor porpoise (*Phocoena phocoena*). Aquat. Mamm. 40, 232–242.

Leaper, R., Calderan, S. and Cooke, J. (2015). A simulation framework to evaluate the

efficiency of using visual observers to reduce the risk of injury from loud sound sources. Aquat. Mamm. 41, 375–387.

Marshall, L. and Brandt, J. F. (1980). The relationship between loudness and reaction time in normal hearing listeners. Acta Otolaryngol. 90, 244–249.

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.

McKenna, M., Calambokidis, J., Oleson, E., Laist, D. and Goldbogen, J. (2015). Simultaneous tracking of blue whales and large ships demonstrates limited behavioral responses for avoiding collision. Endanger. Species Res. 27, 219–232.

Miller, P. J. O., Kvadsheim, P. H., Lam, F. P. A., Tyack, P. L., Curé, C., DeRuiter, S. L., Kleivane, L., Sivle, L. D., van IJsselmuide, S. P., Visser, F., et al. (2015). First indications that northern bottlenose whales are sensitive to behavioural disturbance from anthropogenic noise. R. Soc. Open Sci. 2, 140484.

Mulsow, J., Schlundt, C. E., Brandt, L. and Finneran, J. J. (2015). Equal latency contours for bottlenose dolphins (*Tursiops truncatus*) and California sea lions (*Zalophus californianus*). J. Acoust. Soc. Am. 138, 2678–2691.

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.

NOAA (2015). DRAFT Guidance for Assessing the Effects of Anthropogenic Sound on Marine Mammal Hearing - Underwater Acoustic Threshold Levels for Onset of Permanent and Temporary Threshold Shifts. Maryland, MD: US Department of Commerce.

Nowacek, D. P., Thorne, L. H., Johnston, D. W. and Tyack, P. L. (2007). Responses of cetaceans to anthropogenic noise. Mamm. Rev. 37, 81–115.

Pfingst, B. E., Hienz, R. and Miller, J. (1975). Reaction-time procedure for measurement of hearing. II. Threshold functions. J. Acoust. Soc. Am. 57, 431–436.

Radford, A. N., Purser, J., Bruintjes, R., Voellmy, I. K., Everley, K. A., Wale, M. A., Holles, S. and Simpson, S. D. (2016). Beyond a simple effect: variable and changing responses to anthropogenic noise. In The Effects of Noise on Aquatic

Life II, Advances in Experimental Medicine and Biology, pp. 901–907.

Richardson, W. J., Greene, C. R., Malme, C. I. and Thomson, D. H. (1995). Marine Mammals and Noise. San Diego, CA: Academic Press.

Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene Jr., C. R., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., et al. (2007). Marine mammal noise exposure criteria: Initial scientific recommendations. Aquat. Mamm. 33, 411–521.

Stebbins, W. C. (1966). Auditory reaction time and the derivation of equal loudness contours for the monkey. J. Exp. Anal. Behav. 9, 135–142.

Stone, C. J. (2015). Marine mammal observations during seismic surveys from 1994-2010. JNCC Report no. 463a.

Tougaard, J., Wright, A. J. and Madsen, P. T. (2015). Cetacean noise criteria revisited in the light of proposed exposure limits for harbour porpoises. Mar. Pollut. Bull. 90, 196–208.

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. (2016a). Assessing the effectiveness of ramp-up during sonar operations using exposure models. In The Effects of Noise on Aquatic Life II, Advances in Experimental Medicine and Biology, pp. 1197–1203.

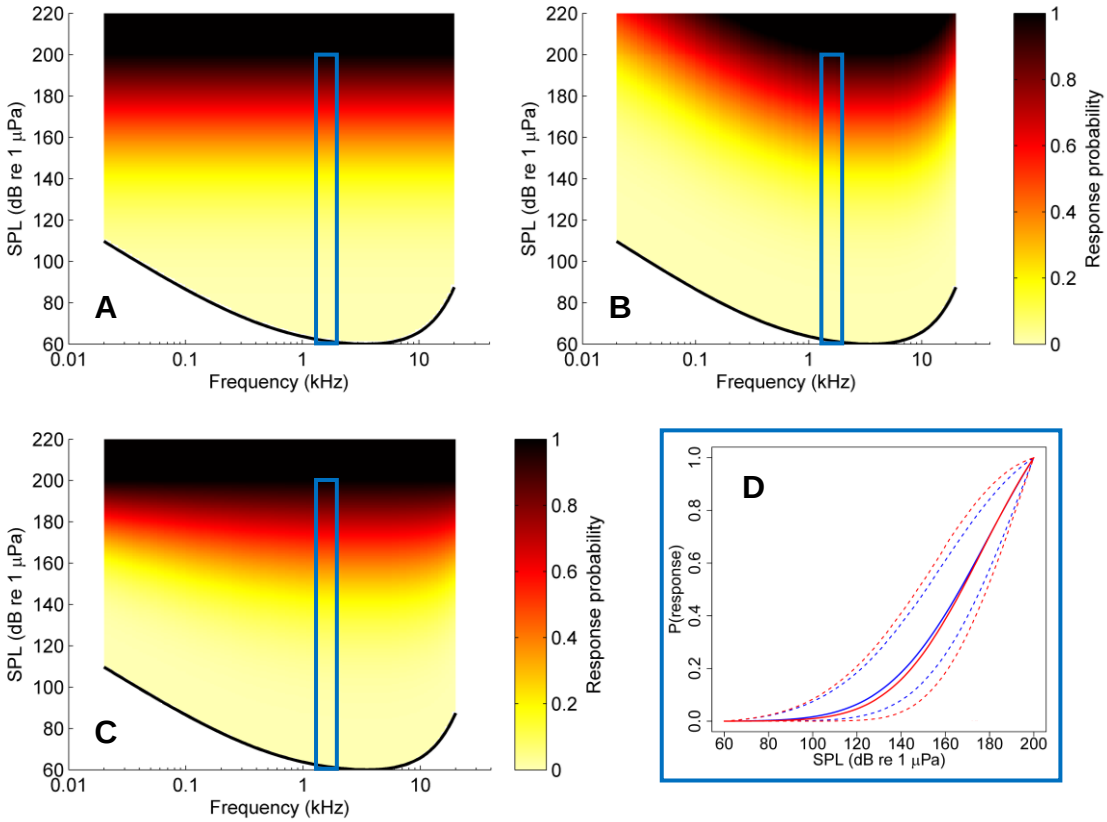
von Benda-Beckmann, A. M., Wensveen, P. J., Samarra, F. I. P., Beerens, S. P. and Miller, P. J. O. (2016b). Separating underwater ambient noise from flow noise recorded on stereo acoustic tags attached to marine mammals. J. Exp. Biol. In review.

Wagner, E., Florentine, M., Buus, S. and McCormack, J. (2004). Spectral loudness summation and simple reaction time. J. Acoust. Soc. Am. 116, 1681–1686.

Wartzok, D., Erbe, C., Getz, W. M. and Thomas, J. (2012). Marine mammal acoustics exposure analysis models used in US Navy Environmental Impact Statements. Adv. Exp. Med. Biol. 730, 551–556.

FIGURES

Figure 6.1. Three examples of how probability of response could be distributed across the hearing range of humpback whales. Extrapolations to frequencies outside the tested sonar band (1.3-2 kHz; blue vertical bars) were made by assuming (A) a SPL-response function that was constant across frequency, (B) a sensation level-response function that was constant across frequency or (C) a scaled SPL-response function with a fixed upper limit at 200 dB re 1 µPa. Examples were created by extrapolation of the dose-response function for avoidance onset (panel D; blue line). The shape of the audiogram (black line) is based on the idealised hearing curve for low-frequency cetaceans presented by NOAA (2015); its maximum sensitivity was anchored at 60 dB re 1 µPa based on the expected hearing abilities of the whale under very low ambient noise conditions.



APPENDICES

Appendix I. Full horizontal tracks and movement parameters for each whale

Figures A1-A13. Shown on the left are the full, most probable track (i.e. the posterior means of $\boldsymbol{x}$) and a detailed view of a section of the track. Visual position fixes were derived from ranges that were estimated by eye or measured using a laser range finder (LRF). Information only shown in the detailed view: GPS positions of the observation boat, computed whale track realisations, and most probable whale positions at the times of the fixes (t_j) with their 95% confidence ellipses (Jackson, 1991). Movement parameters of the track are shown in the panels on the right: (from top to bottom) the whale's body pitch and heading angles measured in the Earth frame, the whale's speed-through-water derived from flow noise, the uncorrected velocity of the whale, the posterior mean velocity correction with 95% credibility intervals (CIs), and the depth of the whale (z-axis coordinate of its position). Note that the scale of the depth axis differs per whale.

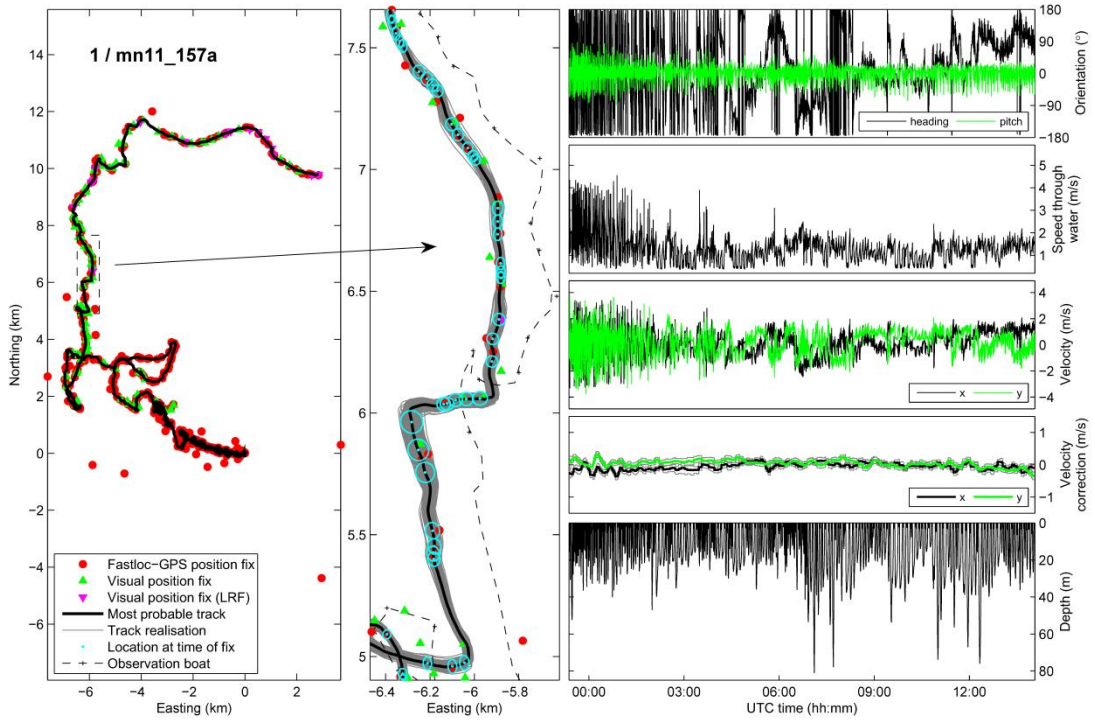


Figure A1.

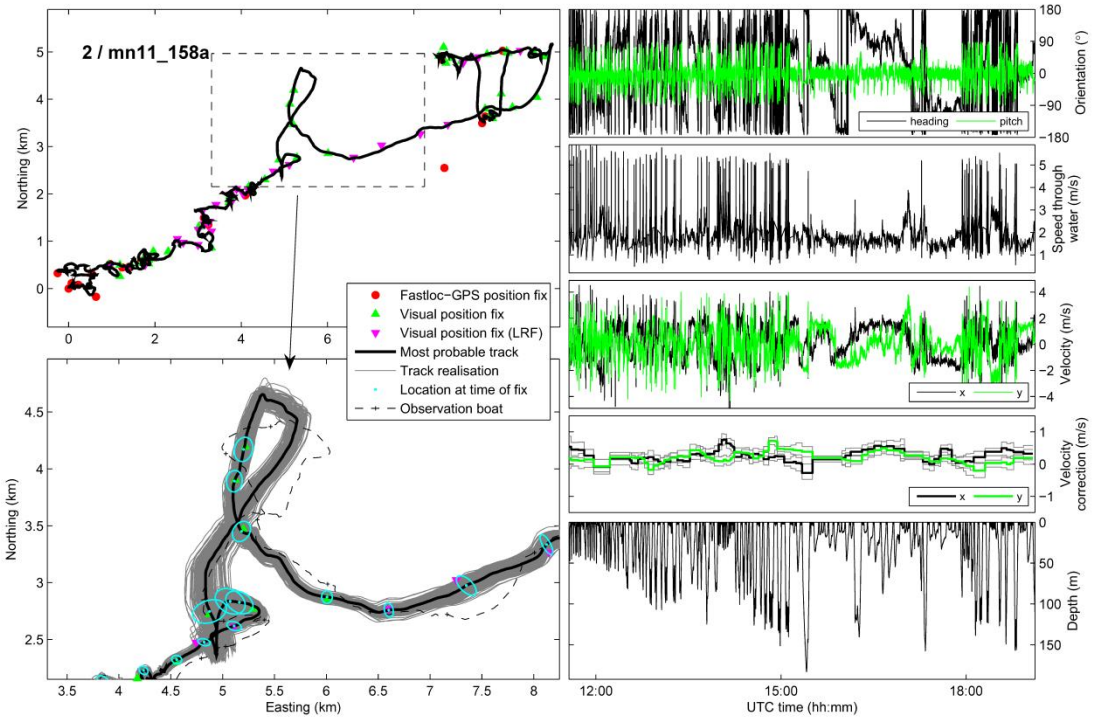


Figure A2.

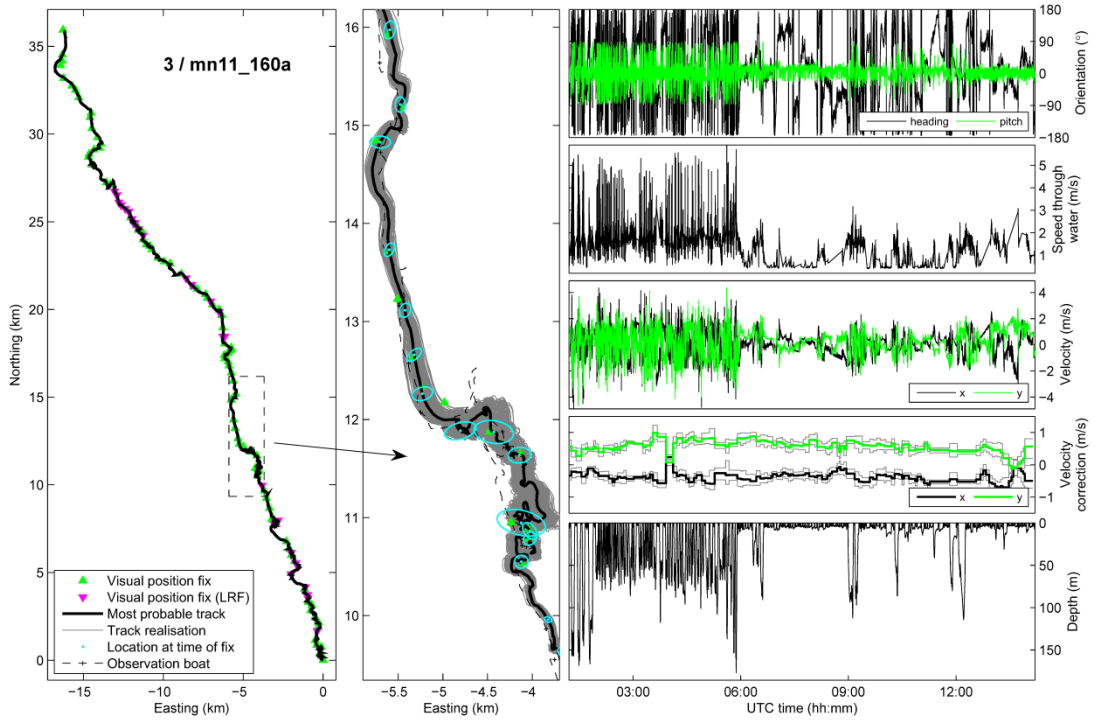


Figure A3.

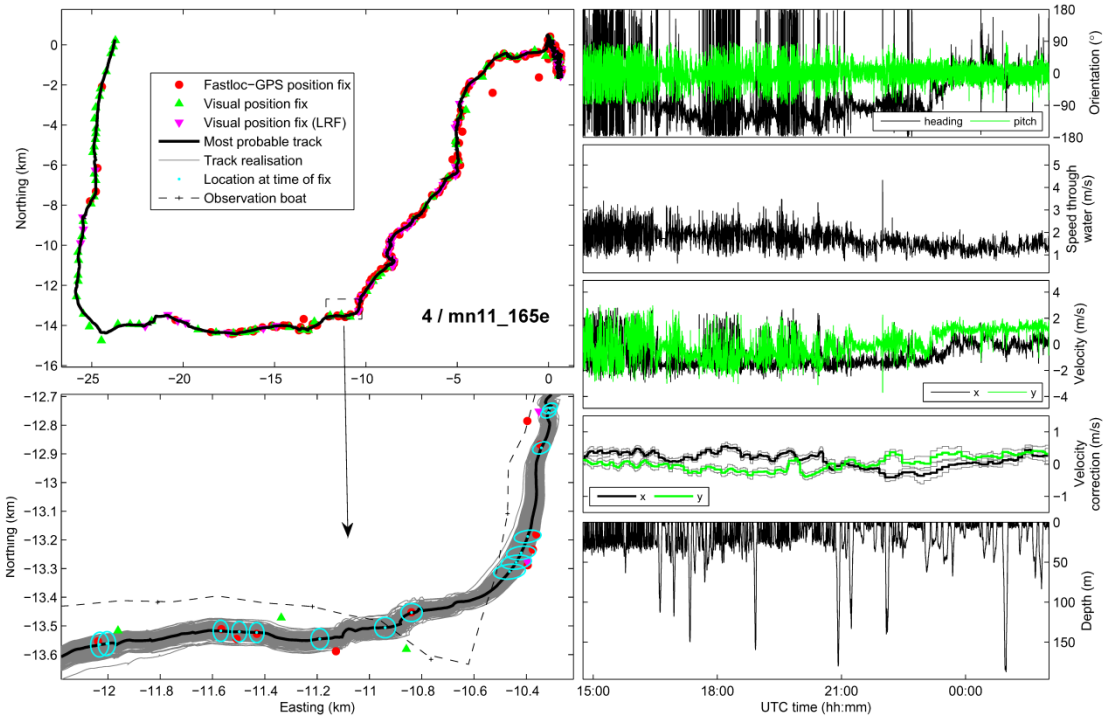


Figure A4.

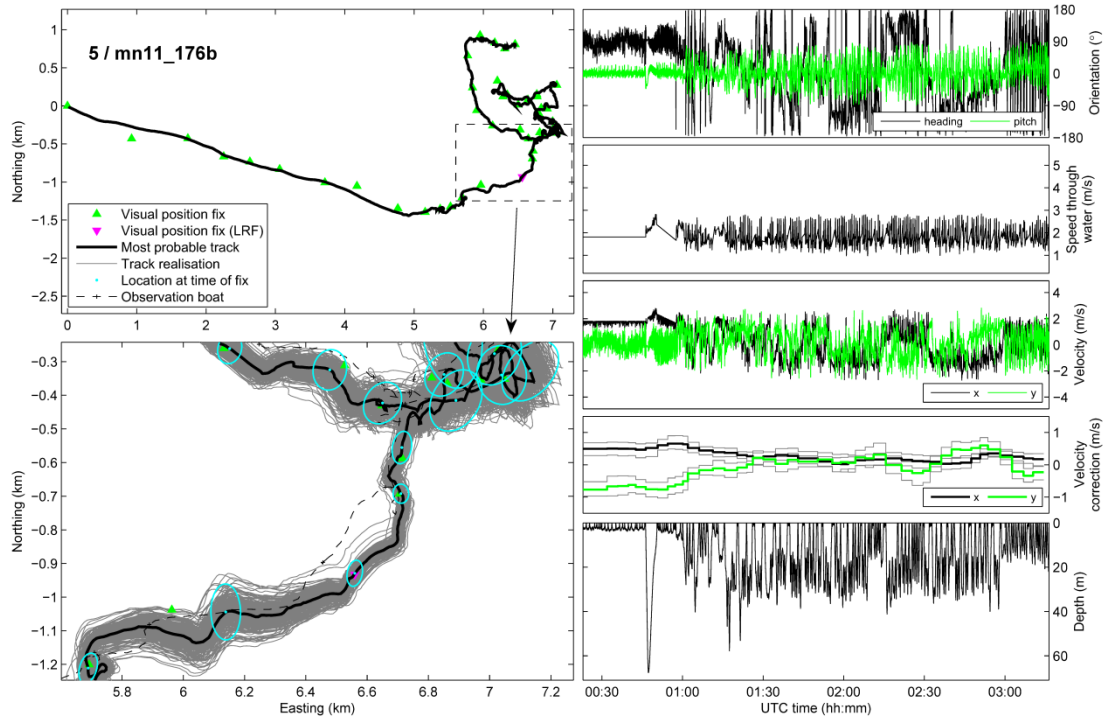


Figure A5.

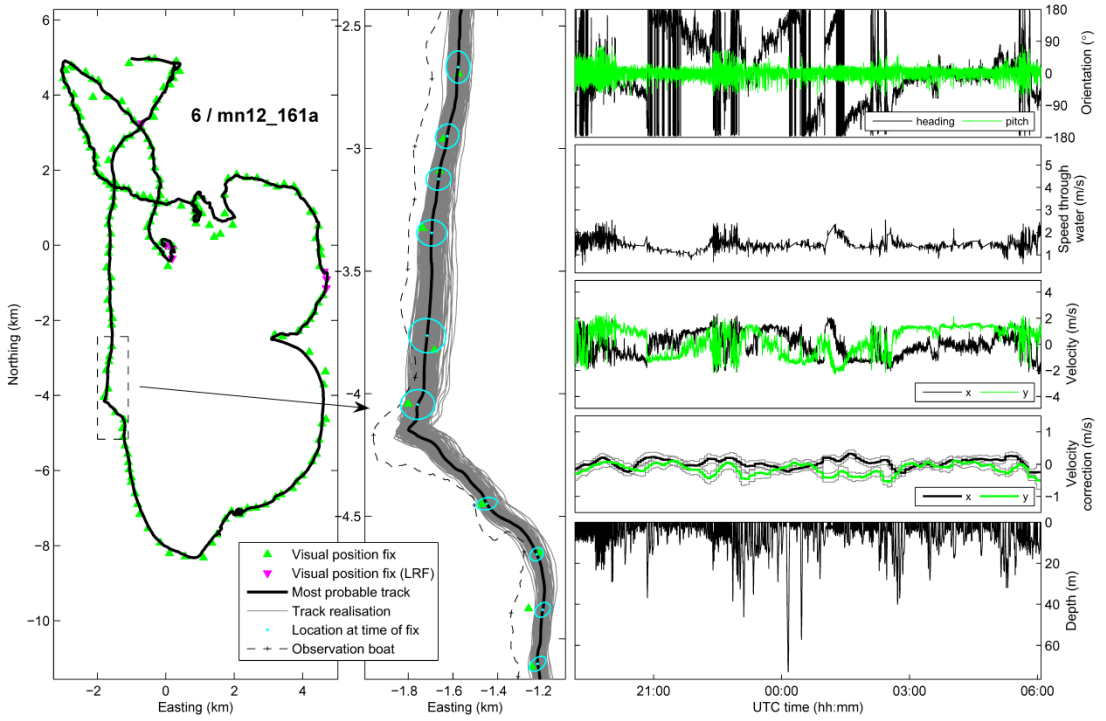


Figure A6.

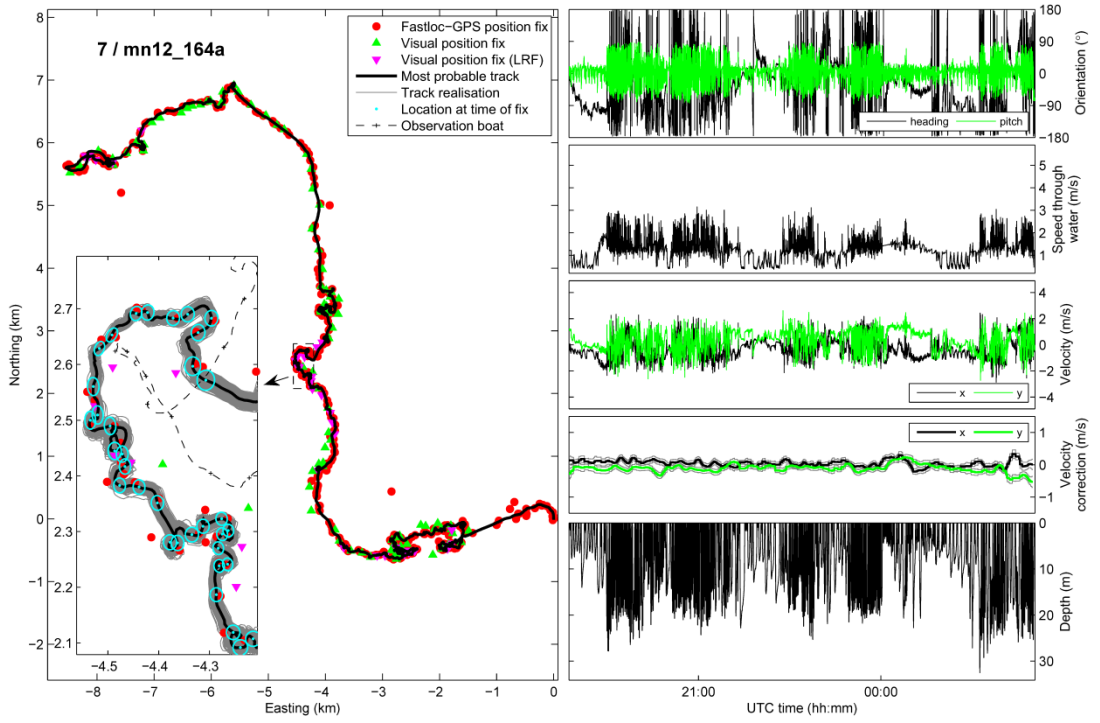


Figure A7.

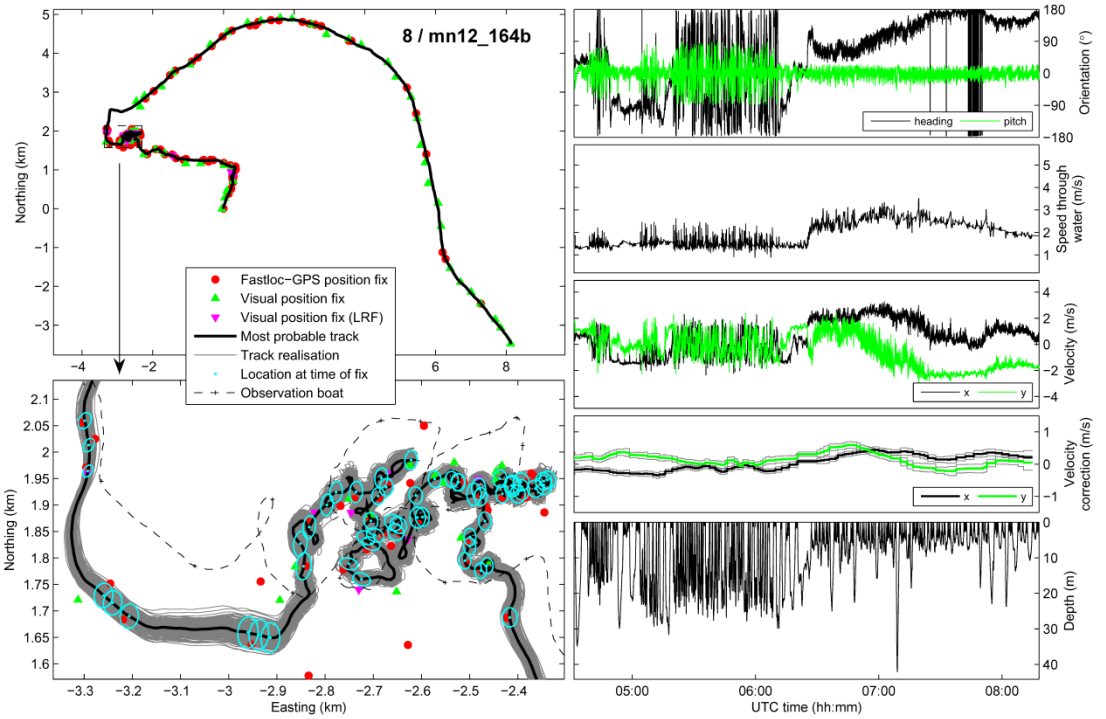


Figure A8.

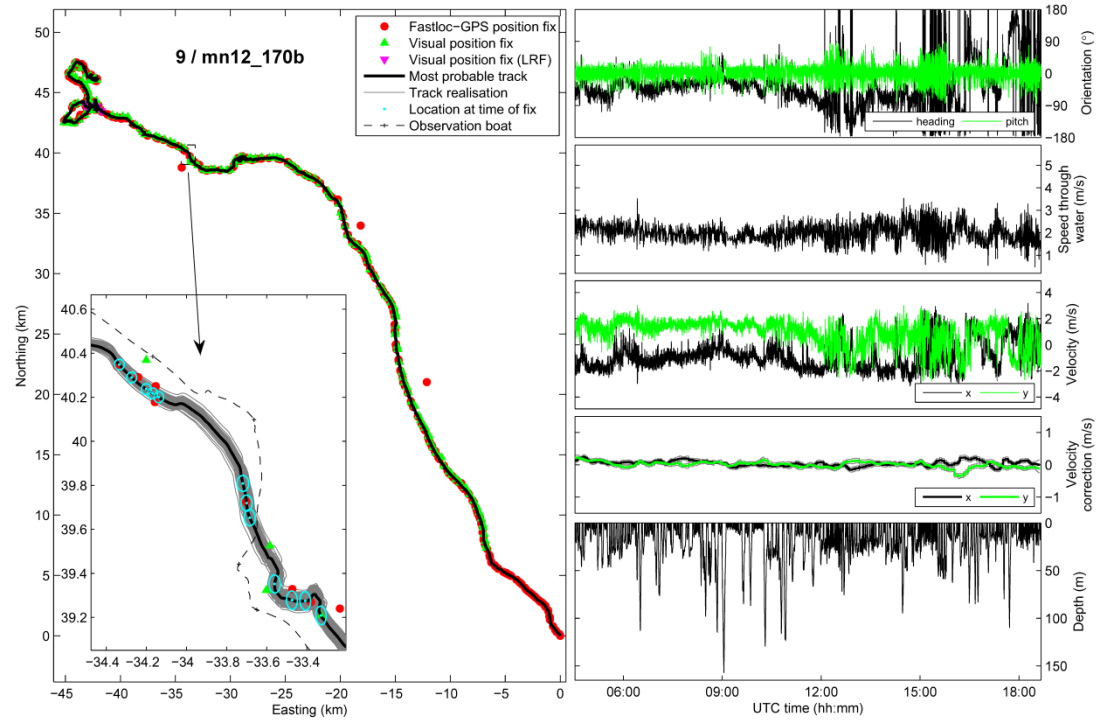


Figure A9.

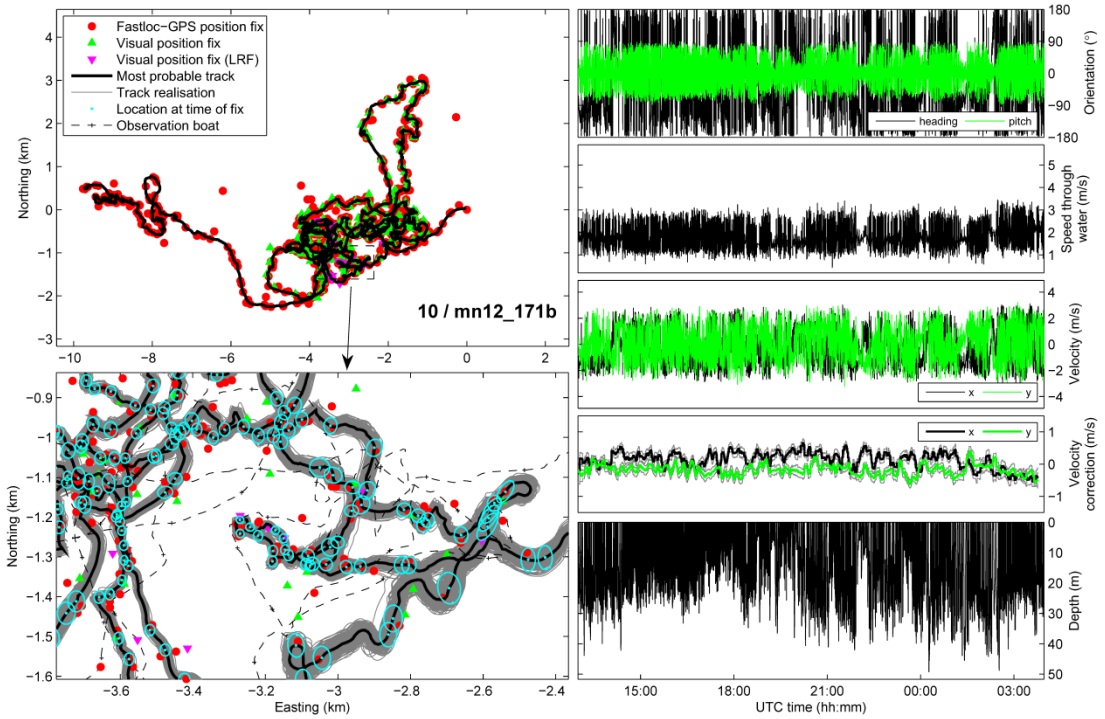


Figure A10.

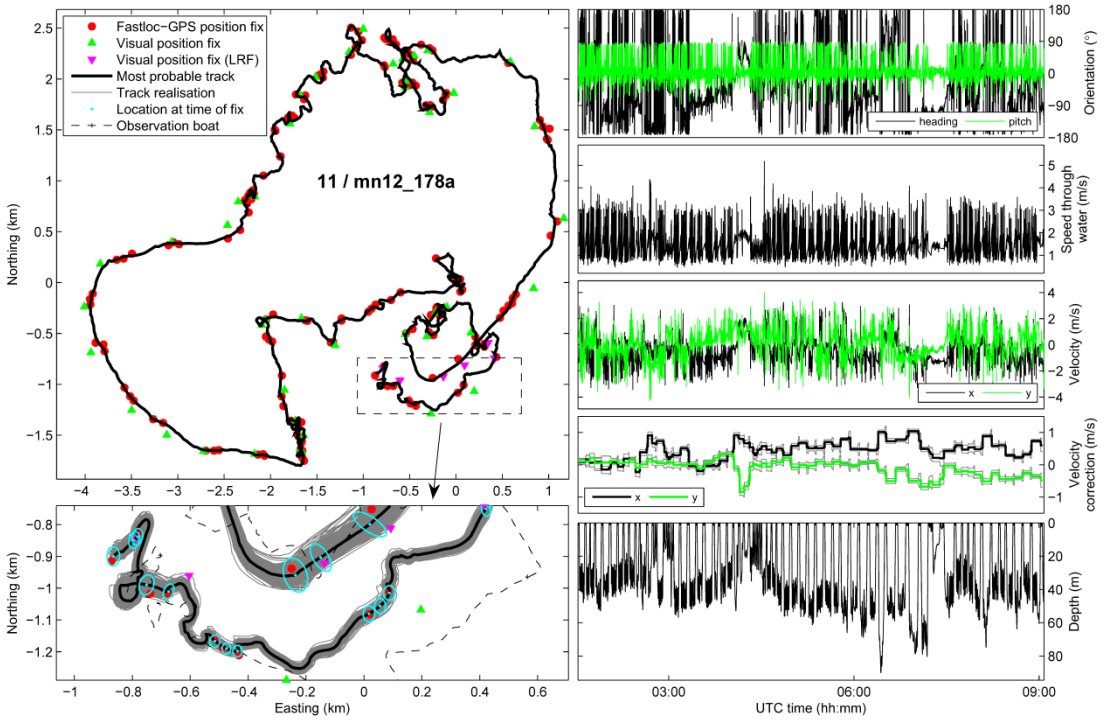


Figure A11.

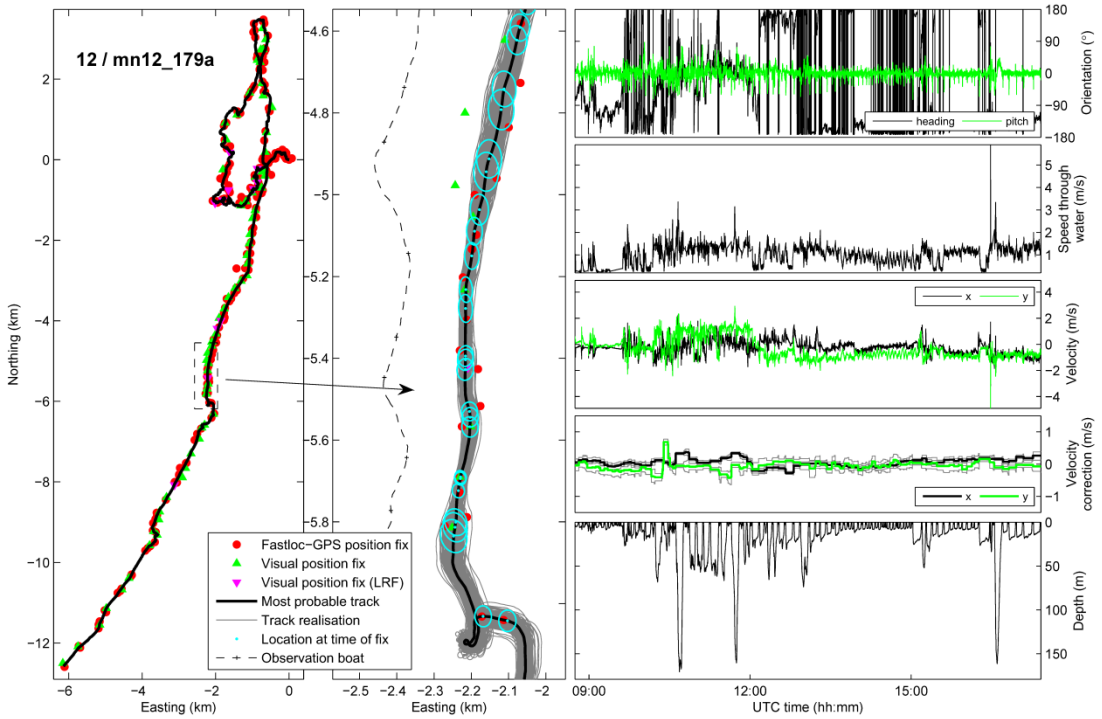


Figure A12.

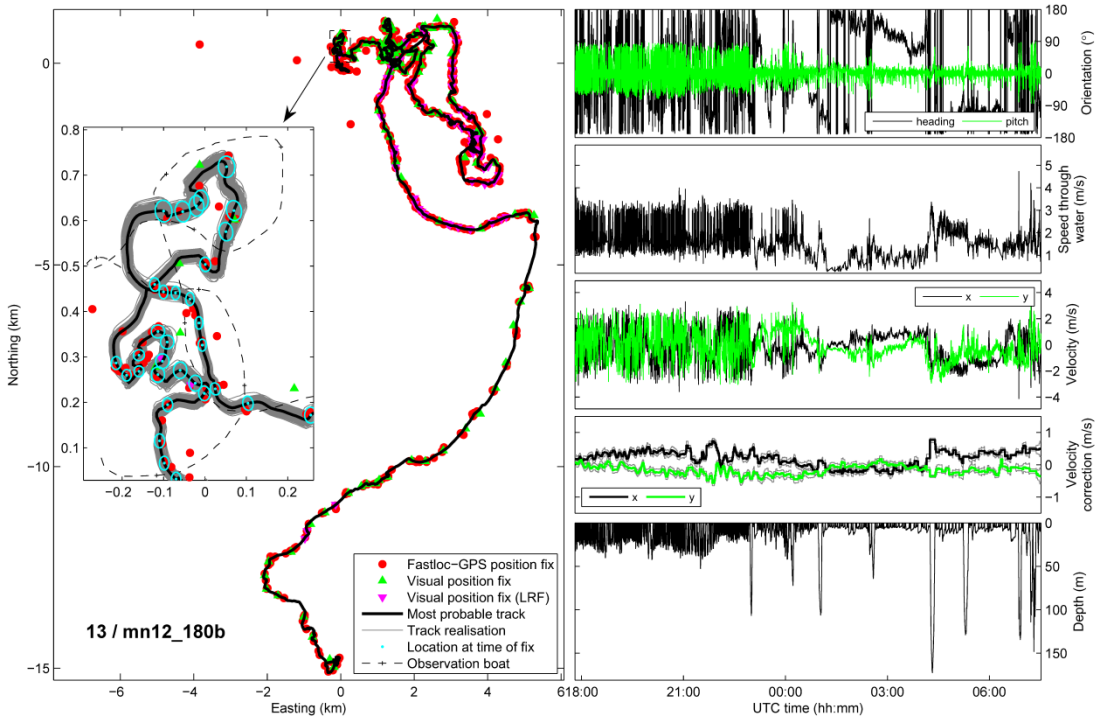


Figure A13.

Appendix II. Horizontal tracks and dive profiles for each exposure session

Figures A14-24. Horizontal movement tracks (top) and time-depth profiles (bottom) of the whales during the different exposure sessions. Positions of the source vessel (black circles) and towed sonar source (black dots) corresponding to individual sonar pulses are shown. Blue, red and green colour-coding indicate whether the whale's movements were from before, during, or after sonar exposure, respectively. The origin (Northing=0; Easting=0) corresponds to the position of the whale at the start of the pre-exposure baseline period. Also indicated is which exposure sessions include a scored behavioural response (Avoidance or Attraction to the source based on this study; or cessation of Feeding based on Sivle et al. 2015). Lunge feeding events of the focal whale are shown with black markers on the dive profile. For sessions with two tagged whales, the horizontal movement tracks and/or time-depth profiles of the non-focal whale are shown with dashed lines; magenta markers on the dive profile indicate feeding events of the non-focal whale. Note that the maximum depth of the panels varies across figures. See Table 3.1 for additional information on each exposure session, such as session type (i.e. *No-Sonar*, *No-RampUp*, or *RampUp*).

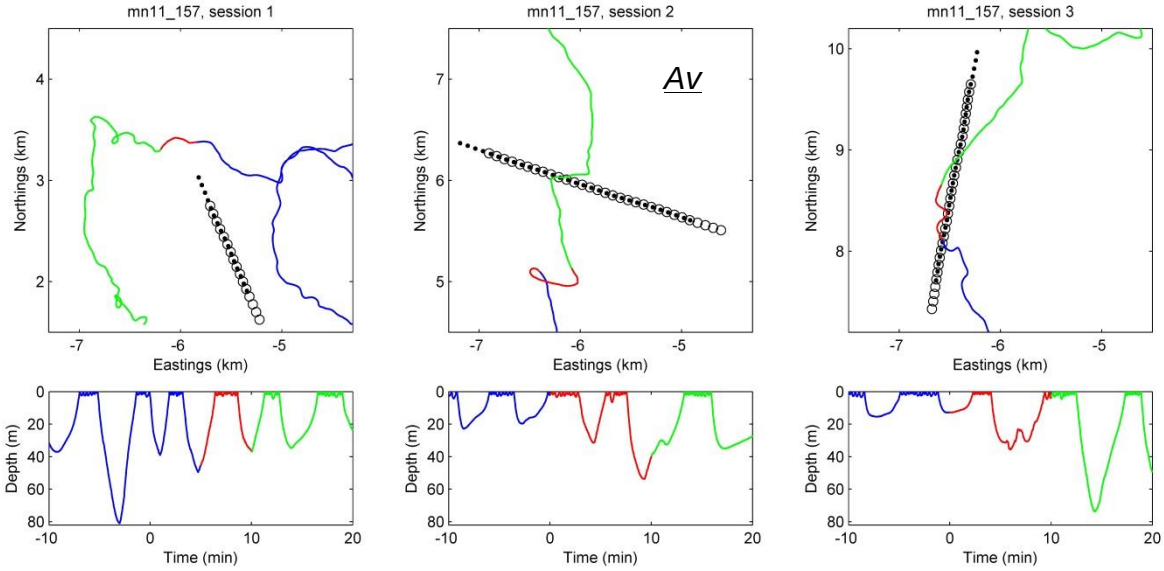


Figure A14. Experiment 1, mn11_157a

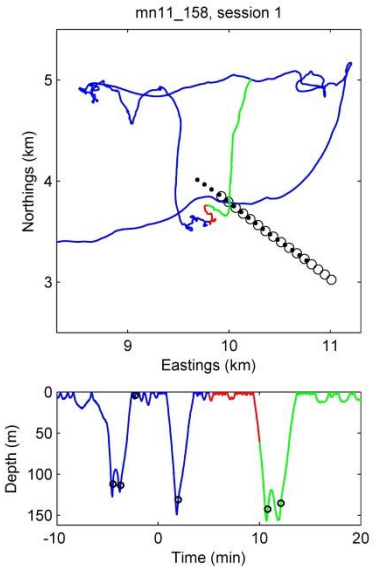


Figure A15. Experiment 2, mn11_158a

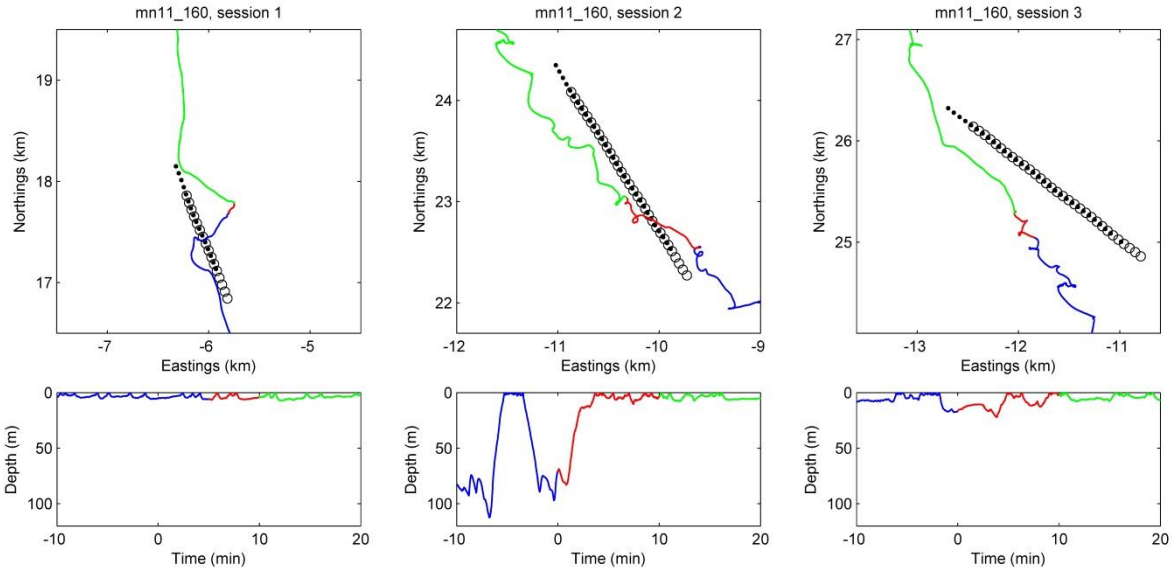


Figure A16. Experiment 3, mn11_160a

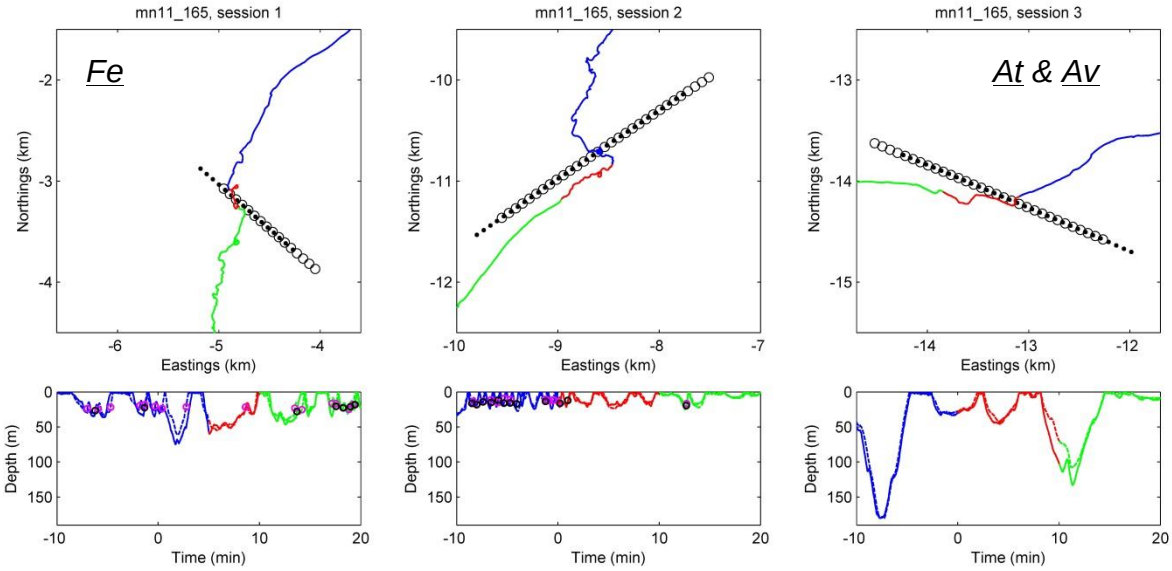


Figure A17. Experiment 4, mn11_165e and f. Tag e was attached to the focal animal. Response in session #1 was scored as "low-confidence" by Sivle et al (2015) because only one animal stopped feeding.

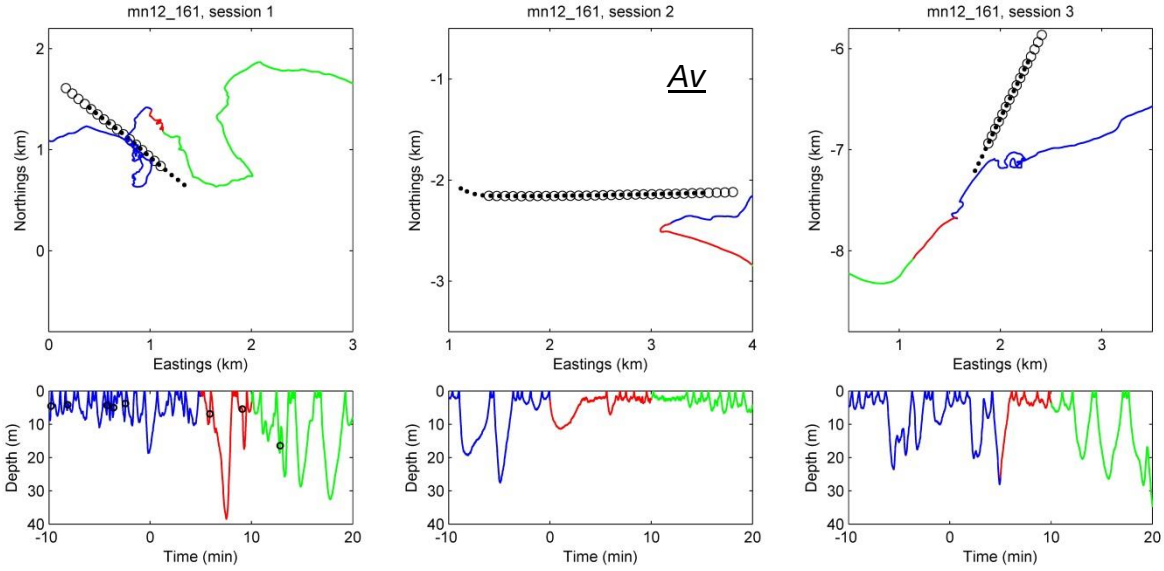


Figure A18. Experiment 5, mn12_161a

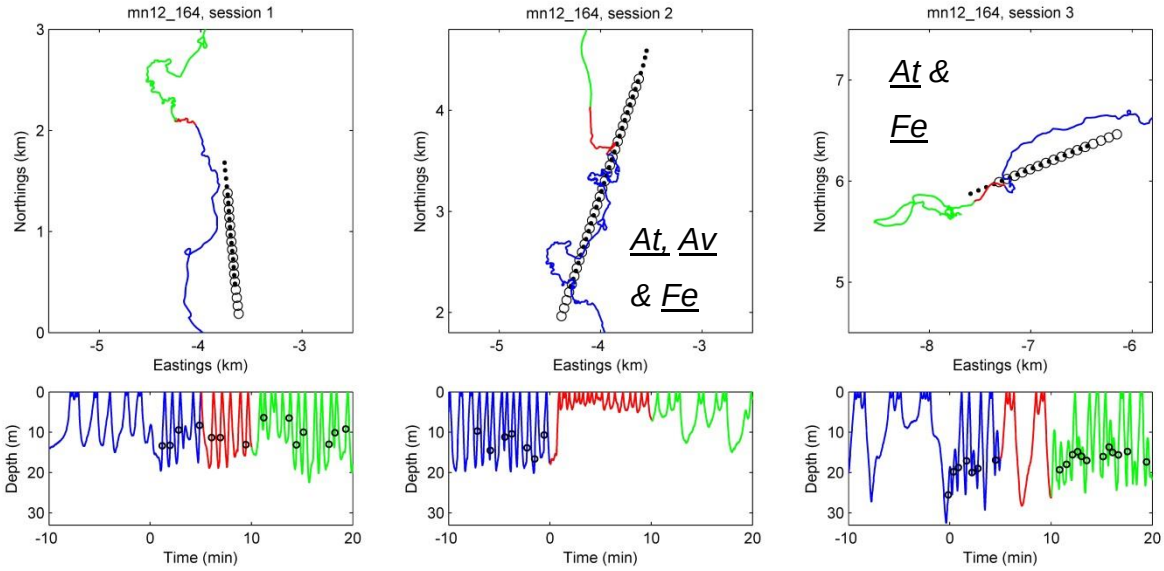


Figure A19. Experiment 6, mn12_164a

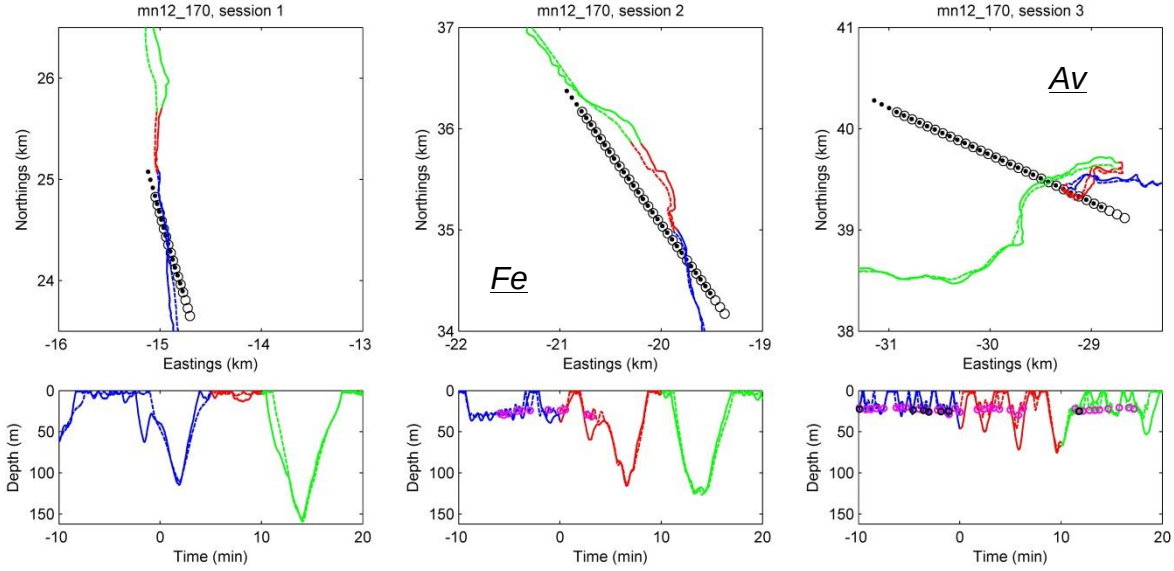


Figure A20. Experiment 7, mn12_170a and b. Tag b was attached to the focal animal.

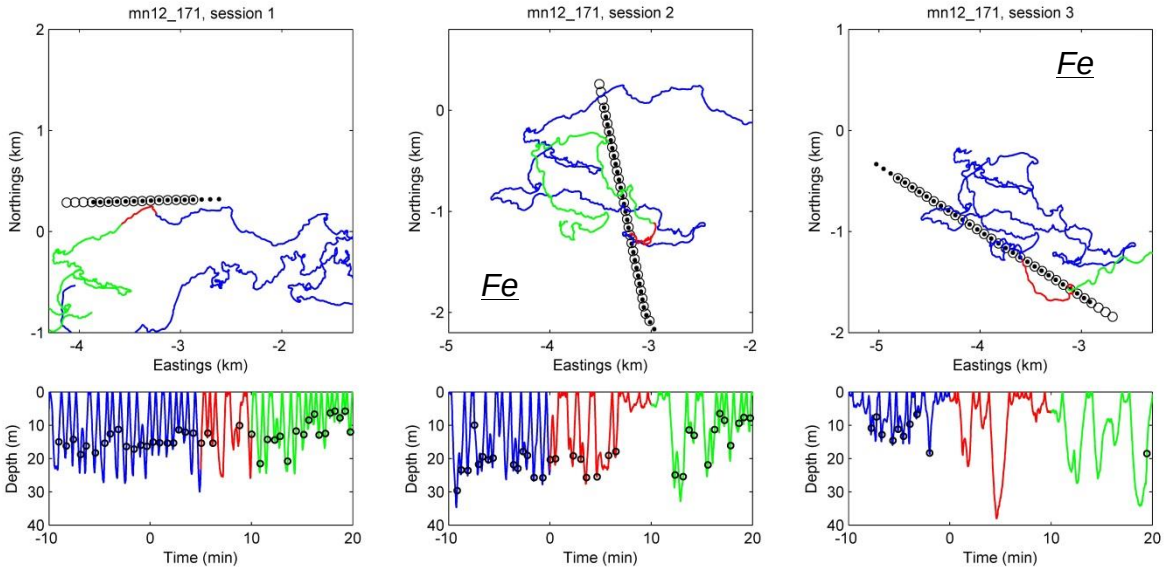


Figure A21. Experiment 8, mn12_171b

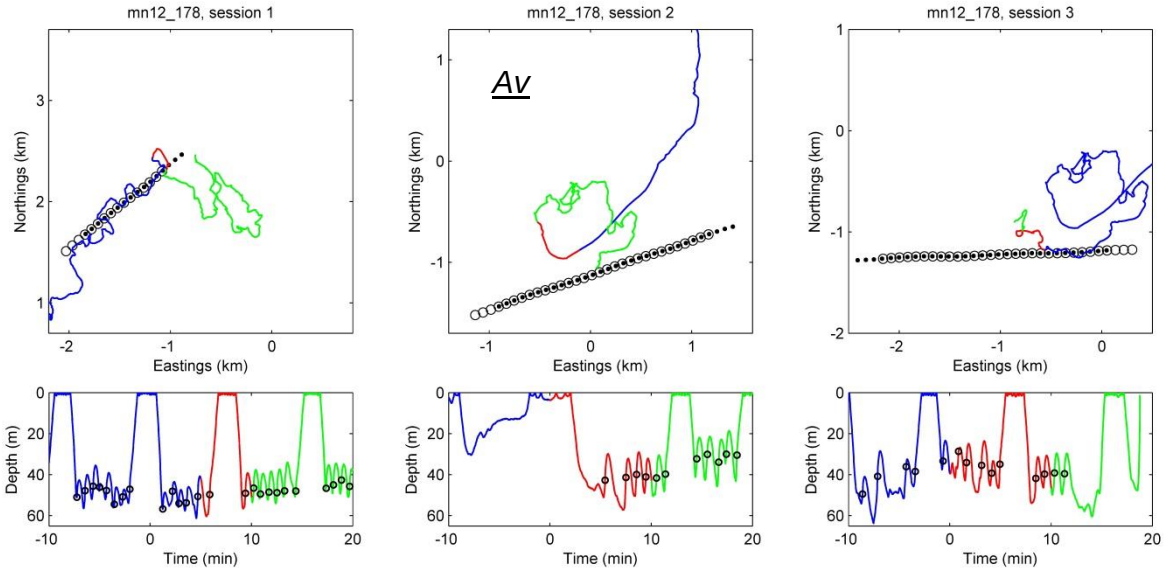


Figure A22. Experiment 9, mn12_178a. The whale breached 9 minutes after the end of session #3, which resulted in a premature tag release.

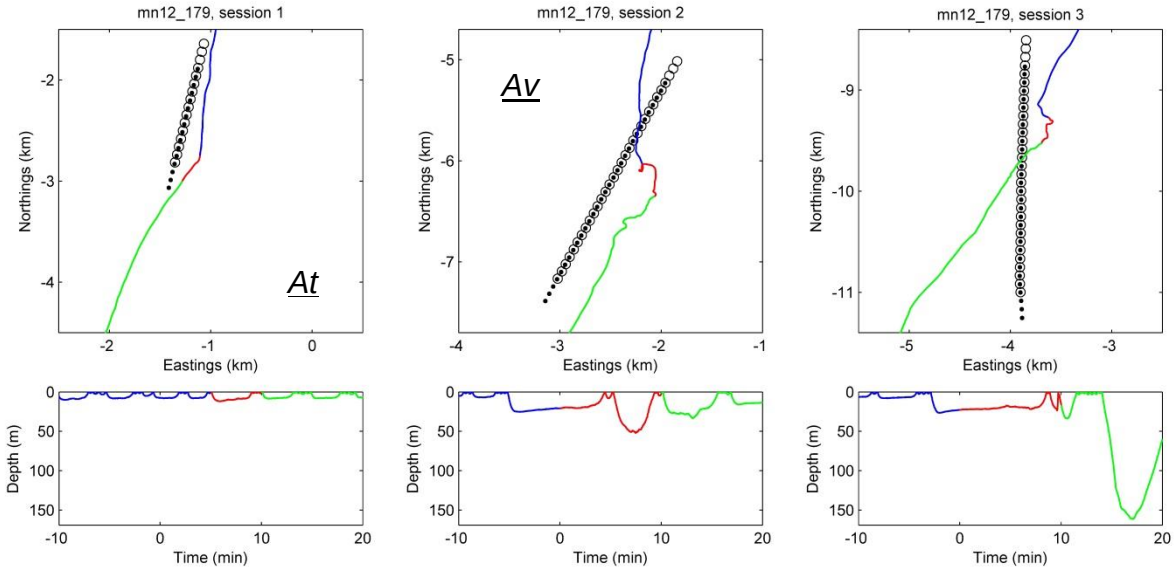


Figure A23. Experiment 10, mn12_179a. The whale breached 9 minutes and 41 s after the start of session #3, 1 s after the penultimate sonar pulse.

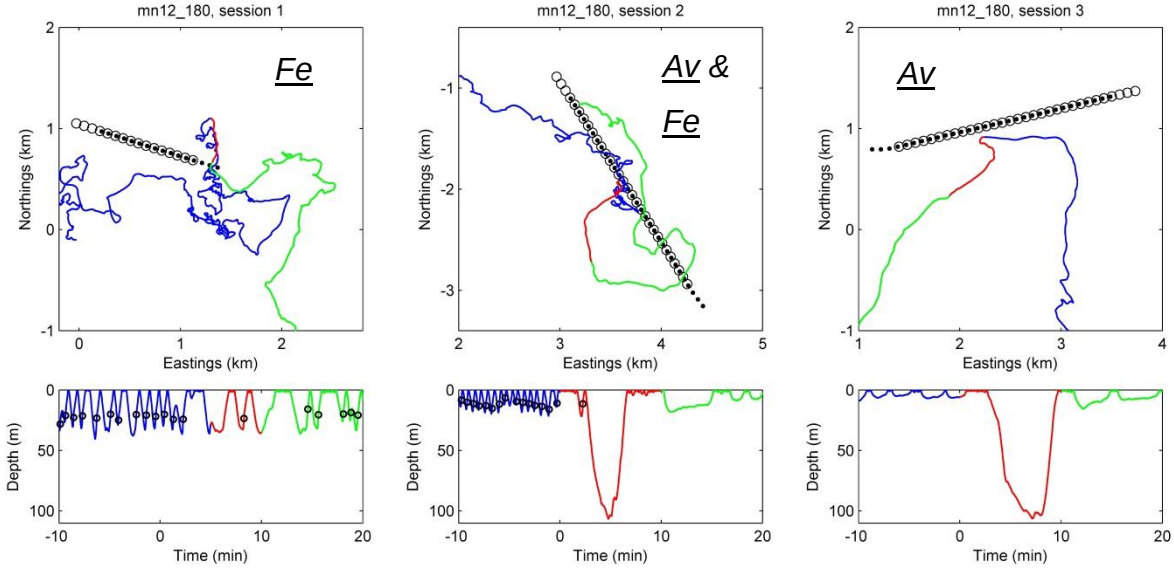


Figure A24. Experiment 11, mn12_180b

Appendix III. Report of the DTAG acoustic calibrations

Measurement report of the DTAG acoustic calibrations in 2011-2013 (tag IDs 235, 237, 238, 241, 242, 243, 246)

Author:

Paul Wensveen (pw234@st-andrews.ac.uk), SMRU

Test Facility:

Anechoic basin of TNO Acoustics and Sonar, The Hague, The Netherlands.

Pool technician: Wim Groen, TNO

Measurement dates:

1 and 4 April 2011

8-9 May 2012

2 May 2013

Introduction

The overall objective of the second phase of the 3S project (3S-2) was to investigate behavioural reactions of cetaceans to low-frequency (1-2 kHz) active sonar signals, in order to establish safety limits for sonar operations. To fulfil this objective, whales were tagged with miniature sound and orientation-recording tags (DTAGs; version 2). The DTAG, developed at WHOI, is a non-invasive tag that can record sound, pressure, as well as 3D accelerometer and 3D magnetometer information. Each tag is equipped with a 16-bit resolution sigma-delta analog-to-digital converter (ADC) with an anti-aliasing filter. The acoustic sensitivity of the DTAG recording chain (hydrophones and signal processing electronics) has to be known to calculate the sound levels of the sonar signals recorded by the DTAGs.

These second-generation DTAGs have either one larger hydrophone to record mono audio data at 96 kHz max ('low-frequency (LF) tags') or two smaller hydrophones to record stereo audio data at 192 kHz max ('high-frequency (HF) tags'). Here we report methods and results of acoustic calibration tests conducted with three LF tags (IDs 237, 238 and 241) and four HF

tags (IDs 235, 242, 243, 246). Calibrations were conducted in the weeks before the annual 3S research trials in 2011, 2012, and 2013.

Methods

Acoustic basin and tag position

Calibration measurements were performed in the anechoic basin (10×8 m² × 8 m deep) at TNO Acoustics and Sonar, The Hague, The Netherlands. The pool contained fresh water during the measurements. In the tank the disadvantage of boundary reverberations for middle and high frequencies are limited as the inside walls and bottom of the tank are covered with an anechoic lining (figure 1).

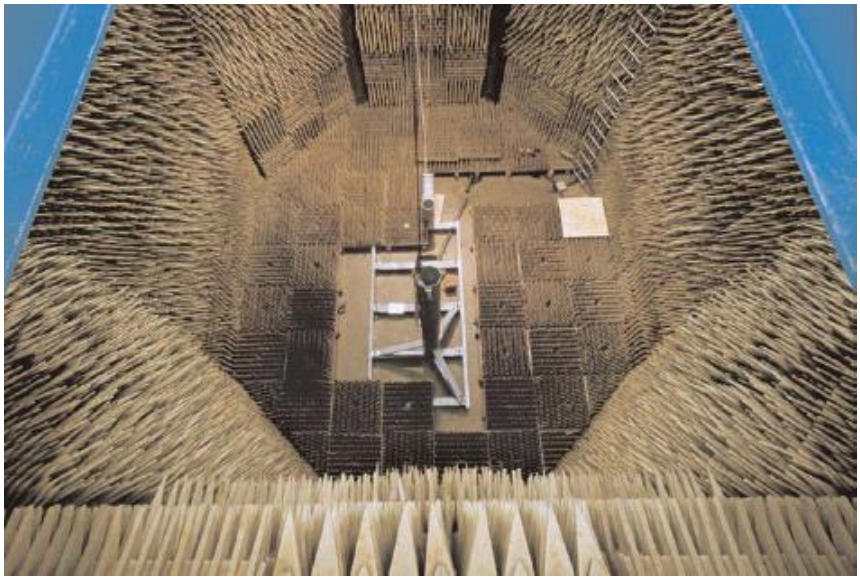


Figure 1. The anechoic basin when it was empty.

The acoustic receiving sensitivity of the DTAGs was determined using a comparison calibration technique. All measurements were made with the tag inside a standard DTAG housing (or, "fairing") with suction cups attached, which also included the standard flotation foam with VHF radio transmitter. Test signals were first recorded using a calibrated reference hydrophone that was positioned at a depth of ~350 cm and at a distance of ~230 cm from the source, in the acoustic axis of the projector. This reference hydrophone was then replaced by the tag, and the DTAG hydrophone(s) were placed at the same position in the pool as the most sensitive part of the reference hydrophone.

Each tag was strapped to a metal rail that was mounted onto a motorised rotating axis. Air bubbles that are trapped inside the housing of the DTAG may influence its acoustic sensitivity. Therefore, before each new series of measurements, 1) the tag was submerged in a bucket of water and shaken to remove the air bubbles, 2) the tag was attached to the metal rail under water with cable ties (figure 2), and 3) the fillet bucket with the tag was lowered into the water and then removed below the pool's water surface, so that the tag stayed submerged and air could not re-enter the housing. A porous polyethylene layer was placed between the tag and the rail to reduce mechanical coupling.

Figure 2. A DTAG submerged in the bucket of water.

The broadband frequency response of the tags was measured with the DTAG hydrophones pointing straight towards the sound source (defined as heading 0°) and the tag positioned horizontally (pitch, roll: 0°). The receiving directivity patterns were measured in the XY and XZ planes by varying the respective heading and pitch angles of the tag (figure 3). Directivity was measured at source angles between 0-360° (XY plane) or 0-180° (XZ plane; from front to top to back) in 20° steps.

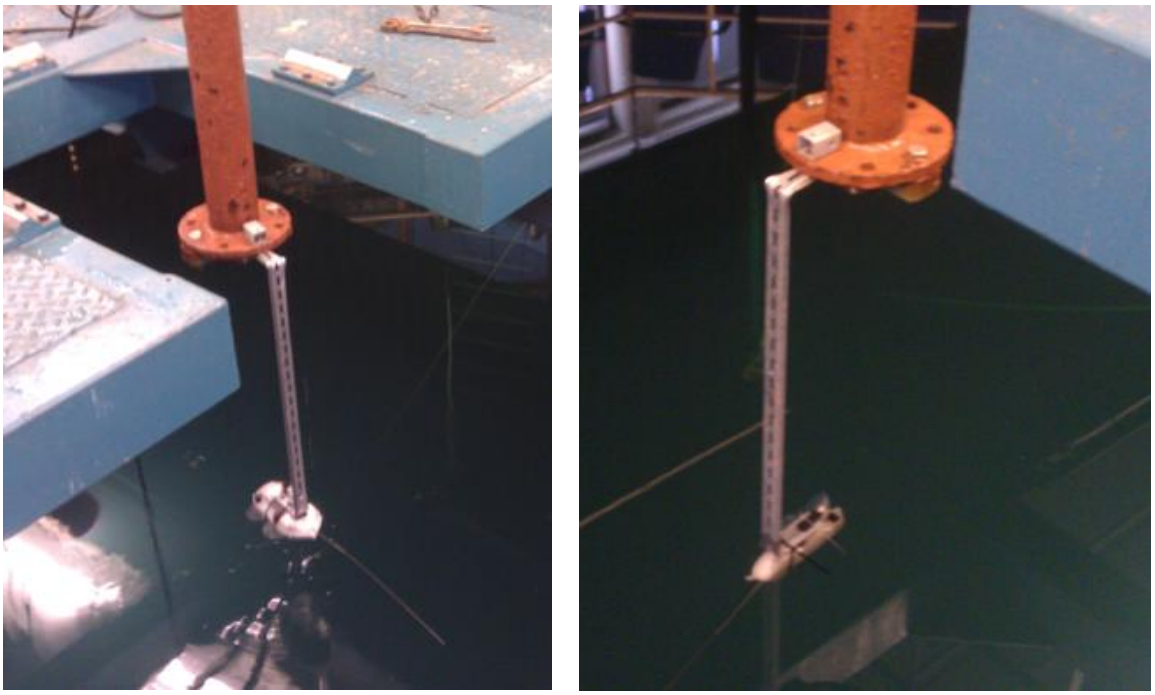


Figure 3. Measurements of the (left) XY-plane and (right) the XZ-plane.

Equipment setup

For details about the transmitting equipment chain (transducer, amplifier, signal generators) and the receiving chain (reference hydrophones, amplifier, data-acquisition system), see table 1 below. The internal gain of the DTAGs was set to 12 dB for LF tags and 0 (channel 1) and 12 dB (channel 2) for HF tags.

Table 1: Equipment and settings used to transmit and record the signals

Transmitting chain	Settings / comments
B&K 1027 sine random generator (for pink noise)	Output 0.5 Vrms, bandwidth 0.02-20 kHz
HP 3314A function generator (for tone bursts, in 2011)	Output 1.42 Vpeak, on manual trigger burst mode, number of cycles was varied to obtain 10 ms
Agilent 33120A function generator (for tone bursts, in 2012-3)	Output 1 Vpp
B&K 2713 power amplifier	Gain 20-26 dB
RANA inductive moving-coil transducer	TNO custom-built moving coil transducer, resembles a USRD J-9

Receiving chain	Settings / comments
B&K 8101 hydrophone (in 2011-2)	s/n 783889 (2011). Calibration chart in figure 7 s/n 783882 (2012). Calibration chart in figure 8
B&K 2610 measuring amplifier (in 2011-2)	Gain 30-40 dB
Reason TC4032-1 hydrophone (in 2013)	s/n 5003146, with built-in pre-amplifier See Table 9 for calibration values
Reason EC6073 input module (in 2013)	
APx521 data-acquisition system	Input range 3.2 Vrms · $2^{(1/2)}$ ~= 4.53 V, sample rate 624 kHz, resolution 24 bits, 2 channels
Dell Precision laptop	APx500 v2.7 or v3.0 software

Acoustic signals and analysis

Broadband pink noise was used as calibration signal to obtain the tag's broadband sensitivity as function of frequency. Digital recordings of the noise were analysed in 1/3-octave bands with centre frequencies between 0.25 and 16 kHz. The RMS amplitude in each 1/3-octave band was obtained by integration of the spectral densities in the band. These spectral densities were calculated from the noise signal by averaging data over time periods of at least 1 min. To obtain the directivity patterns, we used a 10-ms tone burst of 1.5, 6.5 or 15 kHz as the calibration signal. The RMS amplitude of each tone burst was based on a steady portion of the signal that occurred during the first 3.4 ms (= the difference in travel time between the direct path and the shortest indirect path) of the signal. The first cycles of the waveform were excluded so that the amplitude was not influenced by effects related to transducer onset (figure 4).

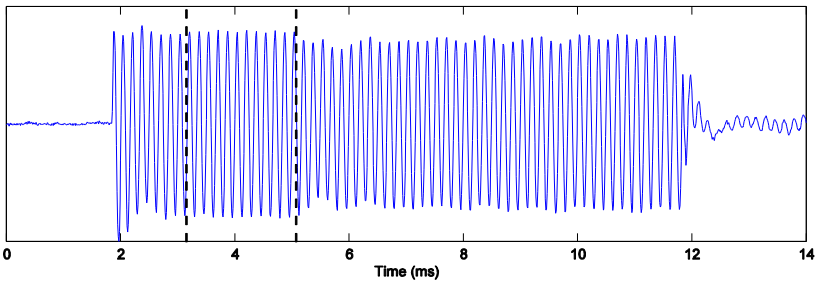


Figure 4. Example of a pulse, with between the dashed lines the selected steady portion.

The sound pressure level (SPL) was calculated from the signals received on the reference hydrophone:

$$SPL = 20\log_{10}(RMS_{ref}) - G_{ref} - S_{daq} - S_{ref}$$

where G_{ref} is the amplifier gain in dB, S_{daq} is the sensitivity of the data-acquisition system in dB re 1/V, and S_{ref} is the voltage sensitivity of the reference hydrophone in dB re V/µPa. The voltage sensitivity of the reference hydrophone for the frequency of the tone burst or centre frequency of the 1/3-octave band was taken from the calibration chart of the hydrophone (appendices A-C).
Because the SPL on the reference hydrophone and DTAG hydrophones is assumed to be identical, the DTAG sensitivity S_{tag} becomes:

$$S_{tag} = 20\log_{10}(RMS_{tag}) - G_{tag} - SPL$$

where G_{tag} is the DTAG internal gain in dB. Reported sensitivity values for the tags correspond to the acoustic sensitivity when the DTAG internal gain is set to 0 dB. The sensitivity of the DTAG is expressed in dB re μPa^{-1} because the reference value of the decibel is expressed in "wavunits" per micropascal.

All measurements were performed by Paul Wensveen (SMRU) and Wim Groen (TNO), with support of Lise Sivle (IMR, Norway) and Filipa Samarra (SMRU) in 2012.

Results: Frequency response curves

The low-frequency DTAGs with one hydrophone were calibrated in all three measurement year (tag 237 and 238) or in the first two years (tag 241). The mean (±SD) sensitivity across the sonar band (1-2 kHz) of tags 237, 238 and 241 was -185.4±2.6, -182.2±1.5, and -181.4±1.6 dB re $1\mu Pa^{-1}$, respectively (figure 5). The frequency response across the whole band (0.25 to 16 kHz) averaged over all three LF tags was flat ±3.1dB, with a mean sensitivity of -183.6 dB re $1\mu Pa^{-1}$.

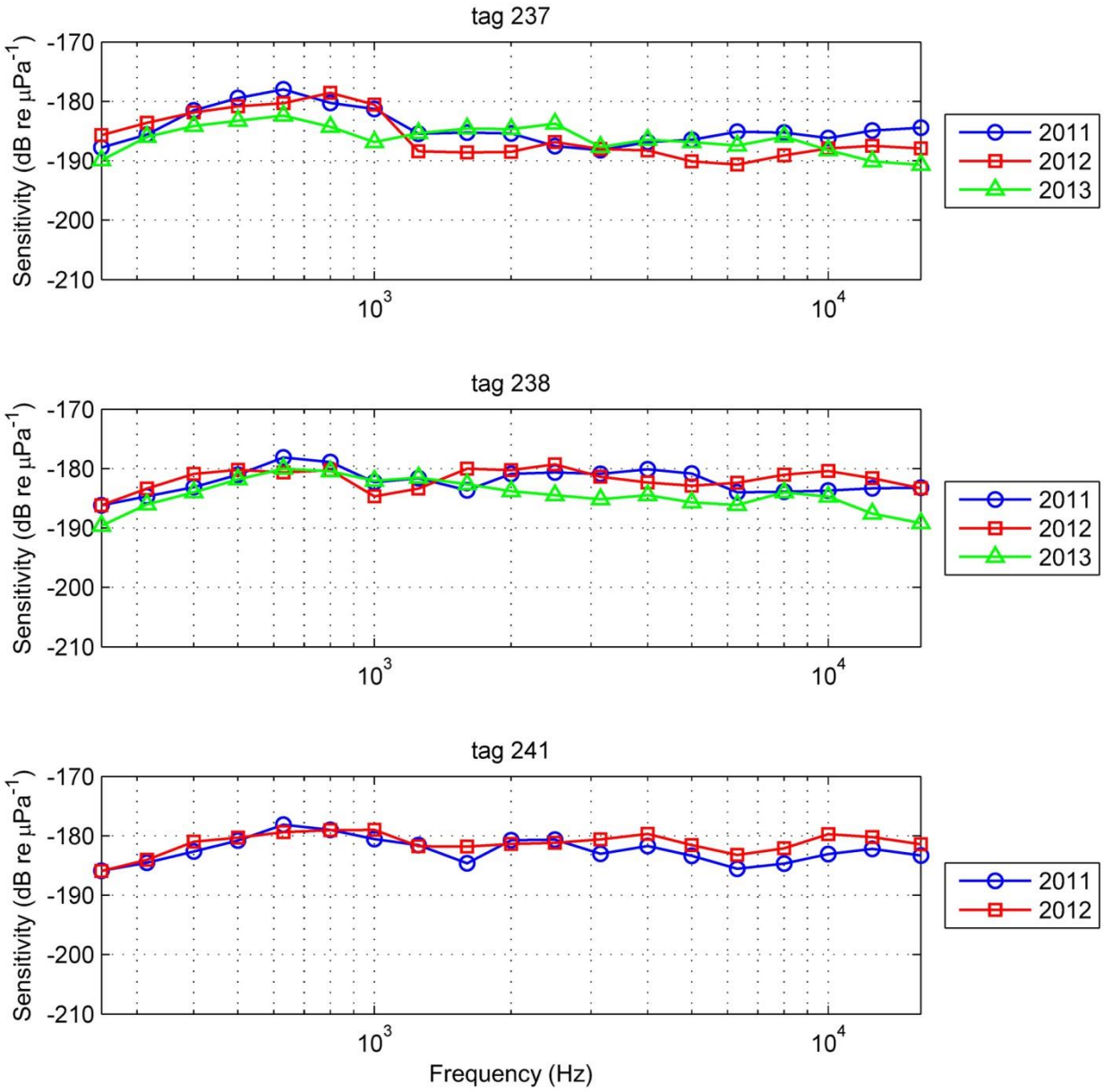


Figure 5. Calibration curves for LF tags. Data for figures can be found in table 3.

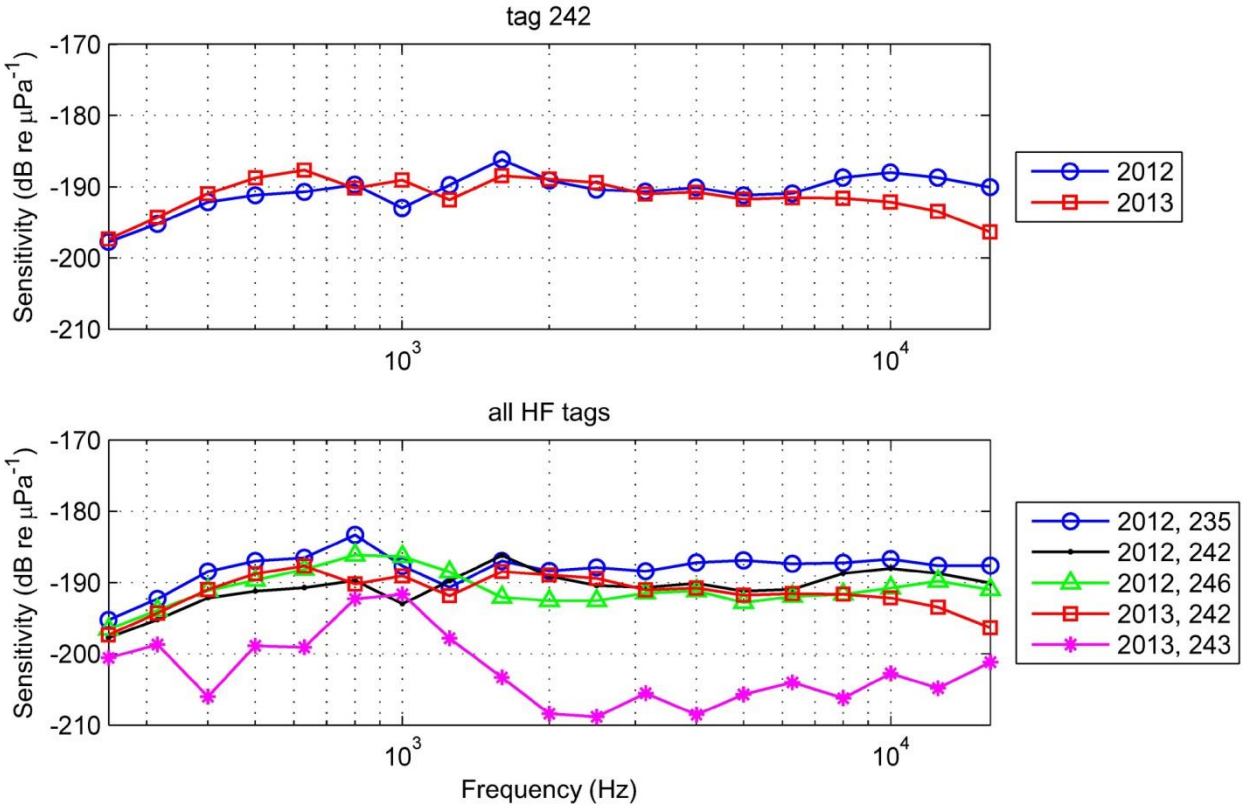


Figure 6. Calibration curves for HF tags. Data for figures can be found in table 3.

The high-frequency DTAGs with two hydrophones were calibrated in two years (tag 242) or only in one year (tags 235, 243 and 246). Figure 6 shows only results for channel 2 (the channel with 12 dB gain) because the spectra of the noise received on channel 1 (no gain) did not exceed the noise floor by enough to provide reliable sensitivity estimates.

Results for tag 242 were comparable between years, except at frequencies above 8 kHz. Its mean (±SD) sensitivity across the sonar band (1-2 kHz) was -189.5±2.1 dB re 1µPa^{-1}. Tags 235, 242 and 246 all had similar sensitivities with, on average, a flat (±2.8dB) frequency response across the whole tested band (0.25 to 16 kHz) and a sensitivity of -190.3 dB re 1µPa^{-1}. Thus, the HF tags were less sensitive than the LF tags by ~6 dB. Overall, these results are consistent with the previous estimate of -188±5dB dB re 1µPa^{-1} for HF tags inside their housing, based on calibration tests of tags 218 and 227 to 230 without housing at TNO and of tags 220 and 227 to 230 with and without housing at NUWC.

Tag 243 was a clear outlier with a reduced sensitivity over the whole frequency band (figure 6). The reason for this is unknown, but the difference may have been caused by the tag itself (hydrophone, pre-amp, ADC) or by air trapped between the tag and the housing.

Receiving directivity patterns

The directivity patterns were generally similar across the 6 tags tested (3 LF tags, 3 HF tags) within each test frequency (tables 2,3,4,5). Table 2 below summarises the results averaged across tag IDs for each type of tag and plane. As expected, the directivity of the tags increased with frequency. At an angle of 180°, directly on the other side of where the hydrophones are placed, sensitivity was 2-3dB higher (1.5 kHz), 4-6dB lower (6.5 kHz), and 14-17dB lower (15 kHz) compared to 0°.

The directivity patterns of single LF tags (tables 6 and 7) show both similarities between years, likely caused by the tag inside the housing, and differences between years, more likely caused by the different housings, trapped air bubbles, or different measurement conditions. Differences due to the position of the two hydrophones in the stereo HF tags were apparent but minor (table 8).

Table 2. Average receiver directivity relative to 0° in the horizontal (left) and vertical (right) plane tested. Data for LF tags are averages over 6 data sets (3 tags, 2 years). Data for HF tags are averages over 3 data sets (3 tags in 2012). Data for figures can be found in tables 4 and 5.

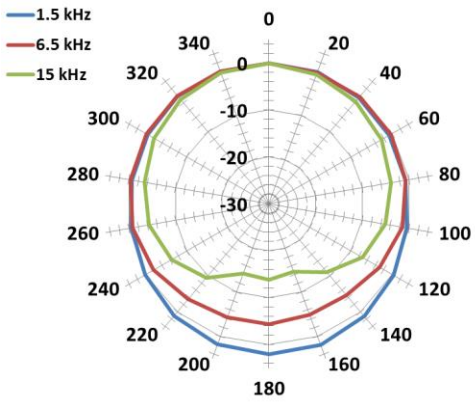
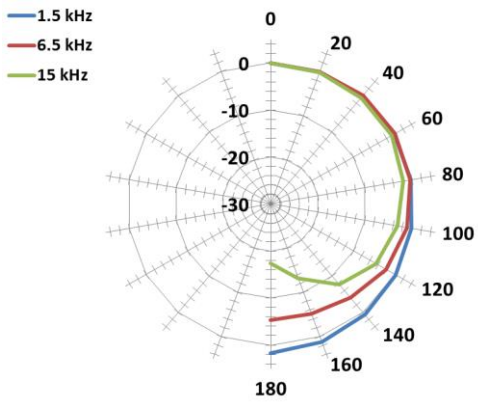
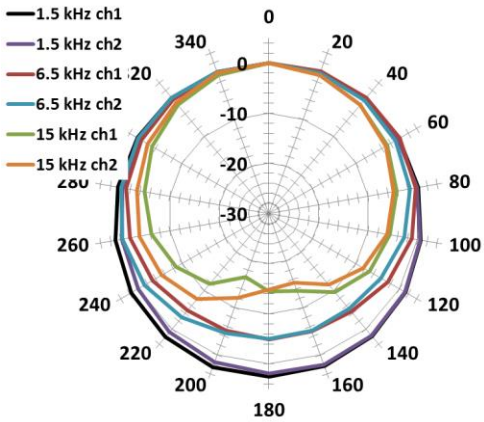
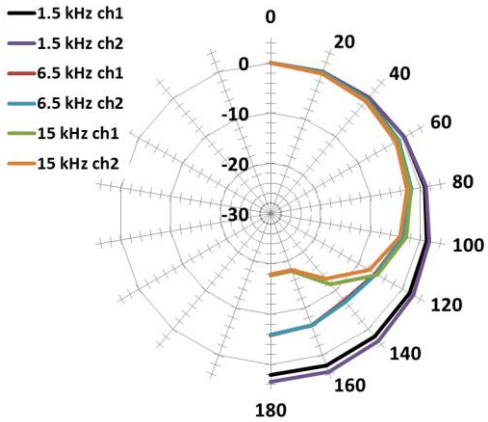
	XY plane (heading)	XZ plane (pitch)
LF tags (IDs 237, 238, 241)		
HF tags (IDs 235, 242, 246)		

Table 3. Sensitivity (dB re 1 μPa^{-1}) of the whole DTAG recording chain (hydrophone, pre-amp, ADC) for all tags when the source was at an angle of 0°. Values correspond to the sensitivity when the DTAG gain is set to 0 dB. Values for channel 1 of the high-frequency tags should not be used (see text).

	2011			**2012**									**2013**					
	Low-frequency tags						**High-frequency tags**						**LF tags**		**High-frequency tags**			
							235		**242**		**246**				**242**		**243**	
f (Hz)	**237**	**238**	**241**	**237**	**238**	**241**	Ch.1	Ch.2	Ch.1	Ch.2	Ch.1	Ch.2	**237**	**238**	Ch.1	Ch.2	Ch.1	Ch.2
250	-187.8	-186.2	-185.9	-185.7	-186.2	-185.9	-193.2	-195.2	-201.3	-197.7	-192.9	-196.5	-189.9	-189.6	-197.4	-197.3	-193.4	-200.6
315	-185.6	-184.7	-184.5	-183.6	-183.3	-184.0	-190.8	-192.3	-200.0	-195.2	-190.3	-193.9	-186.0	-186.0	-194.9	-194.3	-192.4	-198.7
400	-181.5	-183.1	-182.6	-181.9	-180.9	-181.0	-188.1	-188.5	-196.6	-192.1	-188.3	-191.1	-184.1	-184.0	-194.4	-191.0	-189.6	-206.0
500	-179.4	-181.0	-180.8	-180.9	-180.2	-180.3	-187.4	-187.0	-197.8	-191.2	-187.4	-189.7	-183.3	-181.8	-190.9	-188.8	-187.6	-198.9
630	-178.0	-178.1	-178.1	-180.3	-180.6	-179.4	-186.4	-186.6	-197.4	-190.7	-186.1	-188.1	-182.4	-180.0	-188.5	-187.7	-187.4	-199.1
800	-180.3	-178.9	-179.0	-178.5	-180.3	-179.0	-184.8	-183.3	-194.0	-189.7	-185.2	-186.1	-184.3	-180.4	-191.6	-190.2	-187.2	-192.3
1000	-181.3	-182.3	-180.5	-180.5	-184.6	-179.0	-187.6	-187.7	-197.3	-193.0	-184.8	-186.4	-186.8	-182.0	-189.9	-189.1	-187.4	-191.6
1250	-185.4	-181.6	-181.6	-188.4	-183.4	-181.7	-188.5	-190.7	-196.0	-189.8	-186.4	-188.5	-185.3	-181.5	-196.6	-191.8	-186.8	-197.8
1600	-185.3	-183.7	-184.6	-188.6	-180.0	-181.8	-186.6	-186.9	-192.4	-186.2	-189.6	-192.1	-184.6	-182.6	-192.1	-188.5	-186.7	-203.3
2000	-185.4	-180.9	-180.7	-188.5	-180.3	-181.4	-188.8	-188.3	-195.5	-189.1	-190.7	-192.5	-184.7	-183.8	-190.8	-188.9	-186.8	-208.4
2500	-187.5	-180.6	-180.6	-186.8	-179.3	-181.2	-188.6	-187.9	-193.6	-190.4	-192.4	-192.5	-183.7	-184.5	-191.9	-189.4	-189.6	-208.8
3150	-188.2	-180.9	-183.0	-188.0	-181.4	-180.6	-187.5	-188.4	-193.5	-190.7	-191.4	-191.5	-187.7	-185.1	-191.9	-191.0	-193.5	-205.6
4000	-186.9	-180.1	-181.7	-188.2	-182.3	-179.7	-185.9	-187.2	-192.8	-190.1	-190.8	-191.2	-186.4	-184.5	-192.2	-190.8	-193.1	-208.6
5000	-186.5	-180.9	-183.4	-190.1	-182.9	-181.6	-186.1	-186.9	-192.8	-191.2	-191.9	-192.8	-186.8	-185.7	-192.8	-191.8	-195.2	-205.7
6300	-185.1	-184.0	-185.6	-190.6	-182.4	-183.2	-186.9	-187.4	-191.6	-190.9	-191.8	-191.9	-187.4	-186.1	-192.9	-191.5	-196.2	-204.0
8000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	185.2	183.9	184.7	189.1	181.1	182.1	186.6	187.2	189.8	188.7	190.8	191.6	185.9	184.0	194.0	191.6	198.2	206.2
10000	-186.2	-183.7	-183.1	-187.9	-180.4	-179.7	-186.4	-186.7	-189.2	-188.0	-191.1	-190.8	-188.2	-184.7	-194.0	-192.2	-195.7	-202.7
12500	-184.9	-183.3	-182.2	-187.5	-181.6	-180.2	-186.6	-187.6	-189.5	-188.7	-189.6	-189.8	-190.1	-187.6	-194.7	-193.4	-198.9	-204.8
16000	-184.4	-183.2	-183.3	-187.9	-183.3	-181.4	-186.7	-187.6	-191.0	-190.1	-191.2	-191.0	-190.7	-189.2	-196.2	-196.4	-200.2	-201.2

Table 4. Average receiving directivity for in the XY plane for difference heading angles

Angle	**Low-frequency tags**			**High-frequency tags**					
				Channel 1			Channel 2		
deg	dB re 0 deg			dB re 0 deg			dB re 0 deg		
	1.5 kHz	6.5 kHz	15 kHz	1.5 kHz	6.5 kHz	15 kHz	1.5 kHz	6.5 kHz	15 kHz
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	-0.1	-0.2	-0.6	-0.1	0.1	-0.6	-0.2	-0.2	-0.6
40	-0.5	-0.2	-1.5	-0.1	0.1	-1.7	-0.3	-0.5	-1.7
60	-0.7	-0.3	-2.5	-0.1	0.0	-2.8	-0.3	-0.9	-3.2
80	-0.8	-0.7	-3.8	0.2	-0.4	-4.2	0.0	-1.5	-4.8
100	-0.4	-1.5	-5.0	0.6	-1.1	-5.6	0.5	-2.6	-6.1
120	0.4	-3.0	-7.2	1.4	-2.7	-6.9	1.3	-4.2	-8.3
140	1.3	-4.6	-11.1	1.9	-4.5	-9.6	1.8	-5.2	-11.4
160	1.9	-4.8	-14.7	2.4	-5.1	-13.5	2.0	-5.1	-15.4
180	2.1	-4.4	-13.9	2.6	-4.9	-14.5	1.9	-5.0	-14.9
200	1.8	-4.4	-14.3	2.6	-5.2	-16.6	1.5	-4.6	-12.0
220	1.1	-3.5	-9.4	2.2	-4.8	-11.6	0.8	-2.9	-7.8
240	0.2	-1.9	-6.3	1.6	-3.2	-8.7	0.0	-1.4	-5.4
260	-0.5	-0.8	-4.3	1.0	-1.9	-6.3	-0.3	-0.4	-3.8
280	-0.6	-0.3	-3.4	0.5	-1.1	-4.8	-0.4	-0.1	-3.3
300	-0.4	-0.2	-2.0	0.2	-0.7	-3.2	-0.4	0.1	-2.2
320	0.0	0.0	-1.0	0.1	-0.3	-1.8	-0.1	0.1	-1.2
340	0.1	0.0	-0.2	0.1	-0.1	-0.6	0.0	0.1	-0.1

Table 5. Average receiving directivity for in the XZ plane for difference pitch angles

Angle	**Low-frequency tags**			**High-frequency tags**					
				Channel 1			Channel 1		
deg	dB re 0 deg			dB re 0 deg			dB re 0 deg		
	1.5 kHz	6.5 kHz	15 kHz	1.5 kHz	6.5 kHz	15 kHz	1.5 kHz	6.5 kHz	15 kHz
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	0.0	0.0	-0.3	0.1	0.0	-0.3	0.1	-0.1	-0.3
40	-0.1	0.1	-0.5	0.2	-0.2	-0.6	0.2	-0.3	-0.8
60	-0.2	0.0	-0.7	0.6	-0.7	-1.1	0.5	-0.8	-1.4
80	-0.2	-0.4	-1.8	1.0	-1.7	-1.9	1.1	-1.8	-2.6
100	-0.1	-1.0	-3.0	1.3	-3.4	-2.7	1.8	-3.5	-4.0
120	0.2	-2.2	-4.5	1.7	-6.1	-5.7	2.5	-5.9	-7.5
140	0.7	-4.0	-7.7	2.0	-7.4	-11.6	3.1	-7.0	-13.0
160	1.3	-5.2	-13.2	2.2	-6.4	-17.8	3.5	-6.3	-18.1
180	1.7	-5.3	-17.4	2.1	-5.8	-17.6	3.5	-5.9	-17.9

Table 6. Directivity patterns for LF tags in the XY plane (angle is heading)

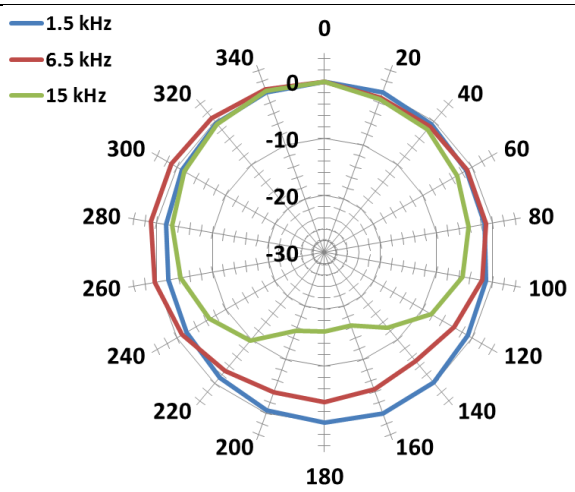
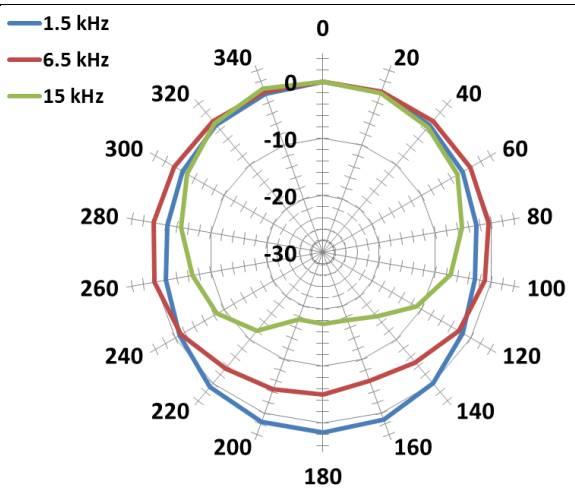
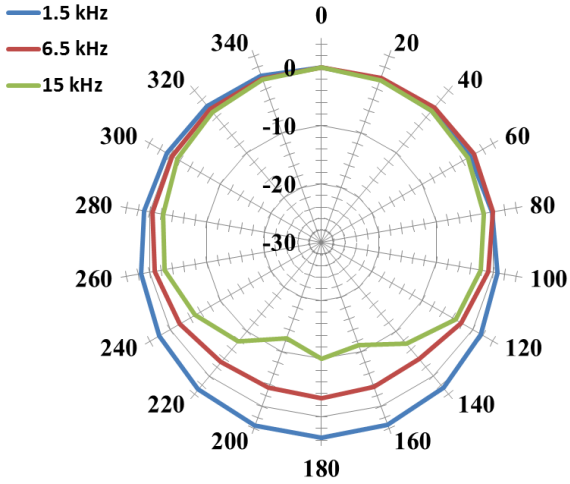
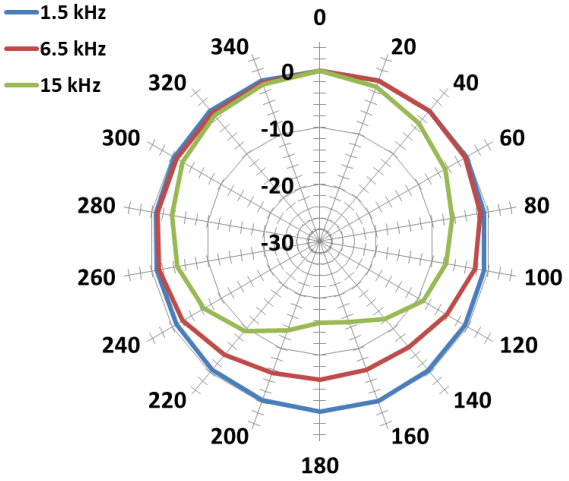
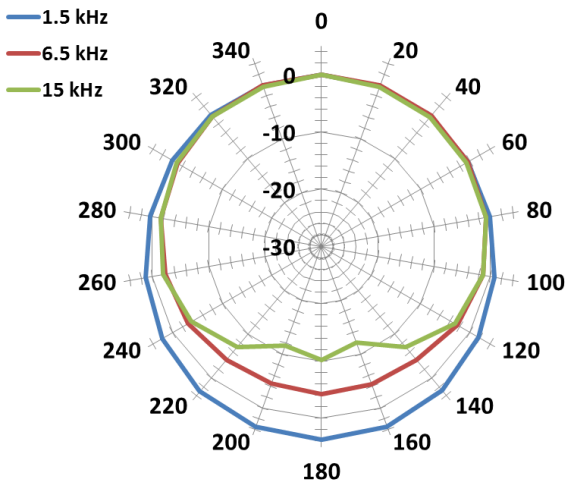
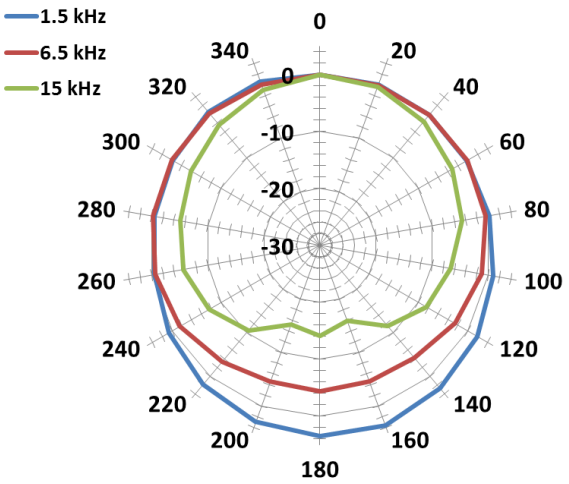
Tag id	2011	2012
237		
238		
241		

Table 7. Directivity patterns for LF tags in the XZ plane (angle is pitch)

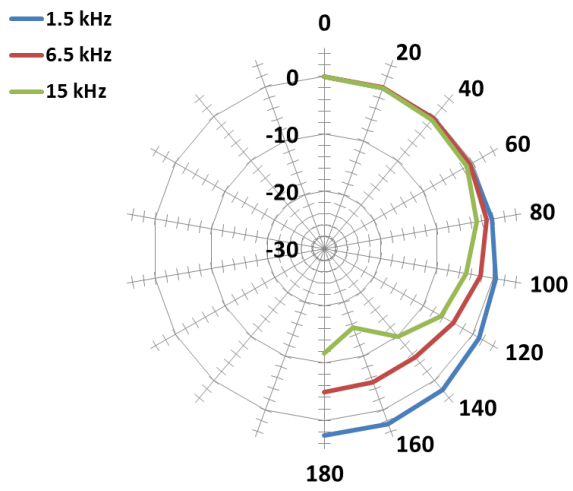
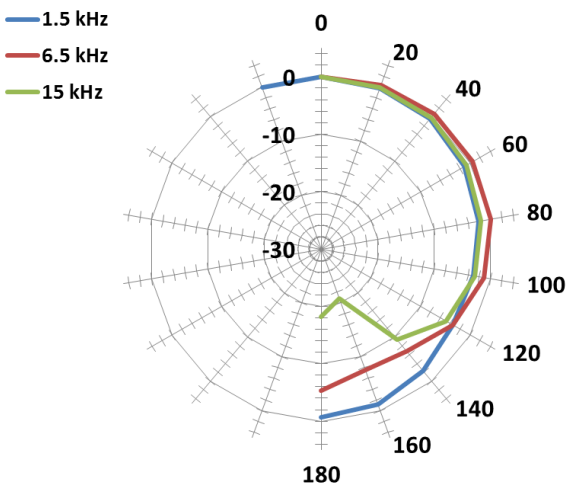
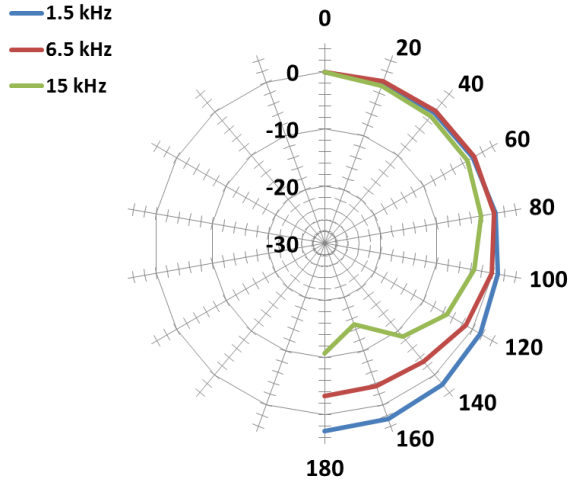
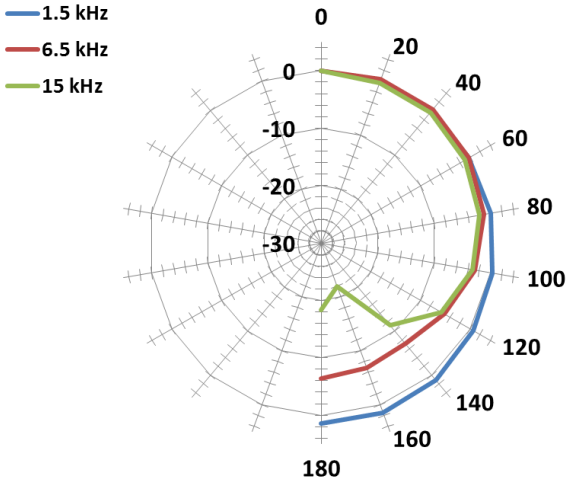
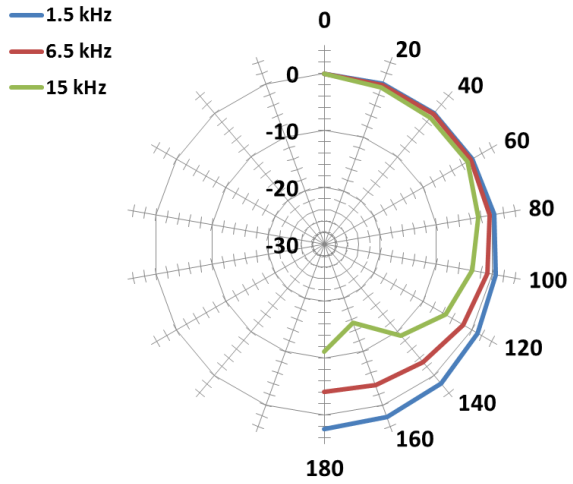
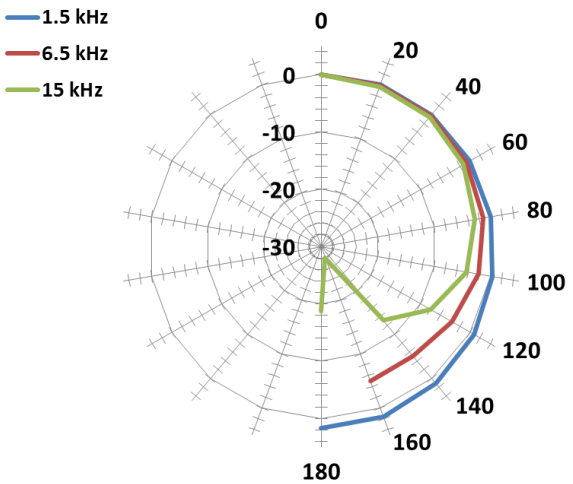
Tag id	2011	2012
237		
238		
241		

Table 8. Directivity patterns for HF tags (only measured in 2012)

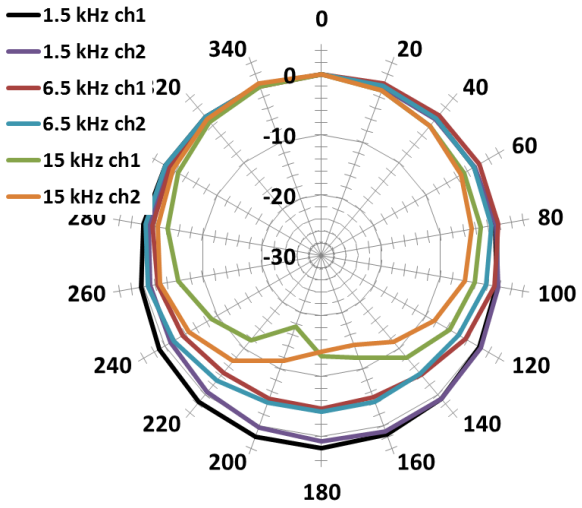
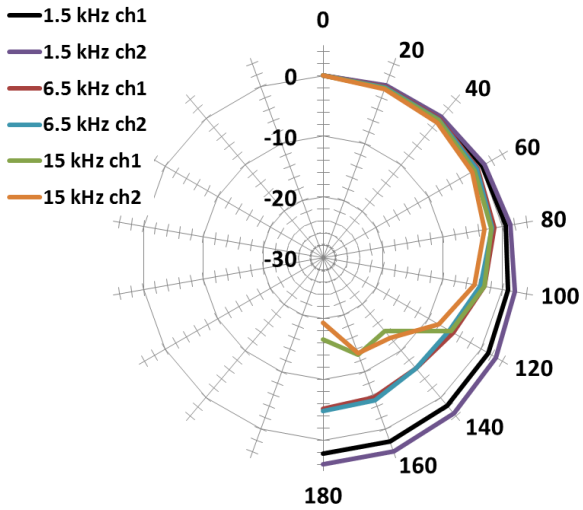
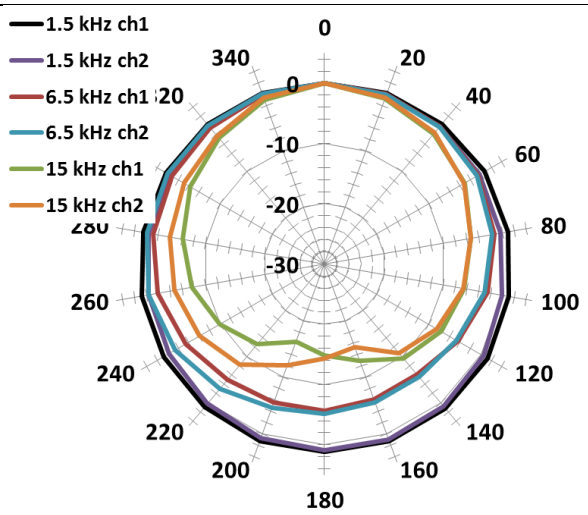
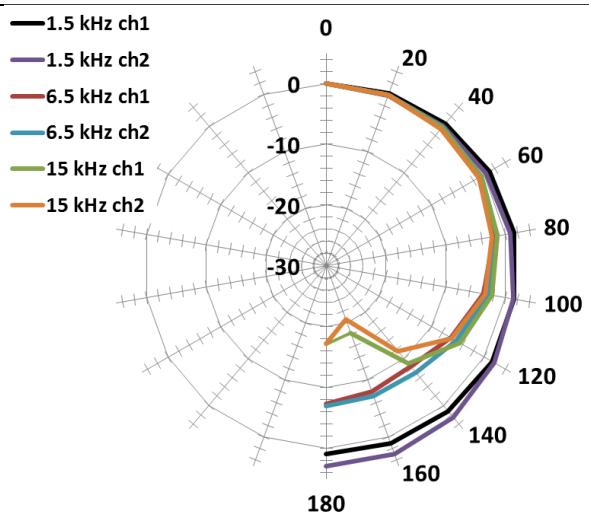
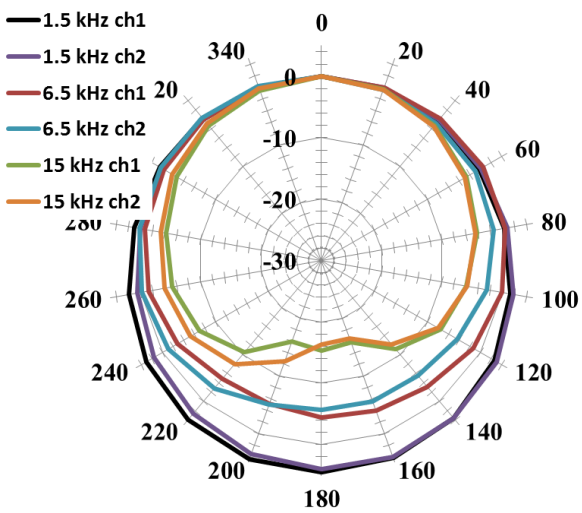
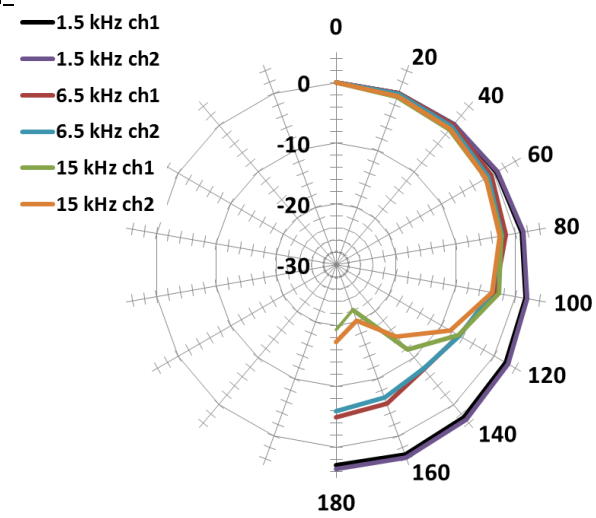
Tag id	XY plane (heading)	XZ plane (pitch)
235		
242		
246		

Fig. 7. Calibration chart for the reference hydrophone used in 2011

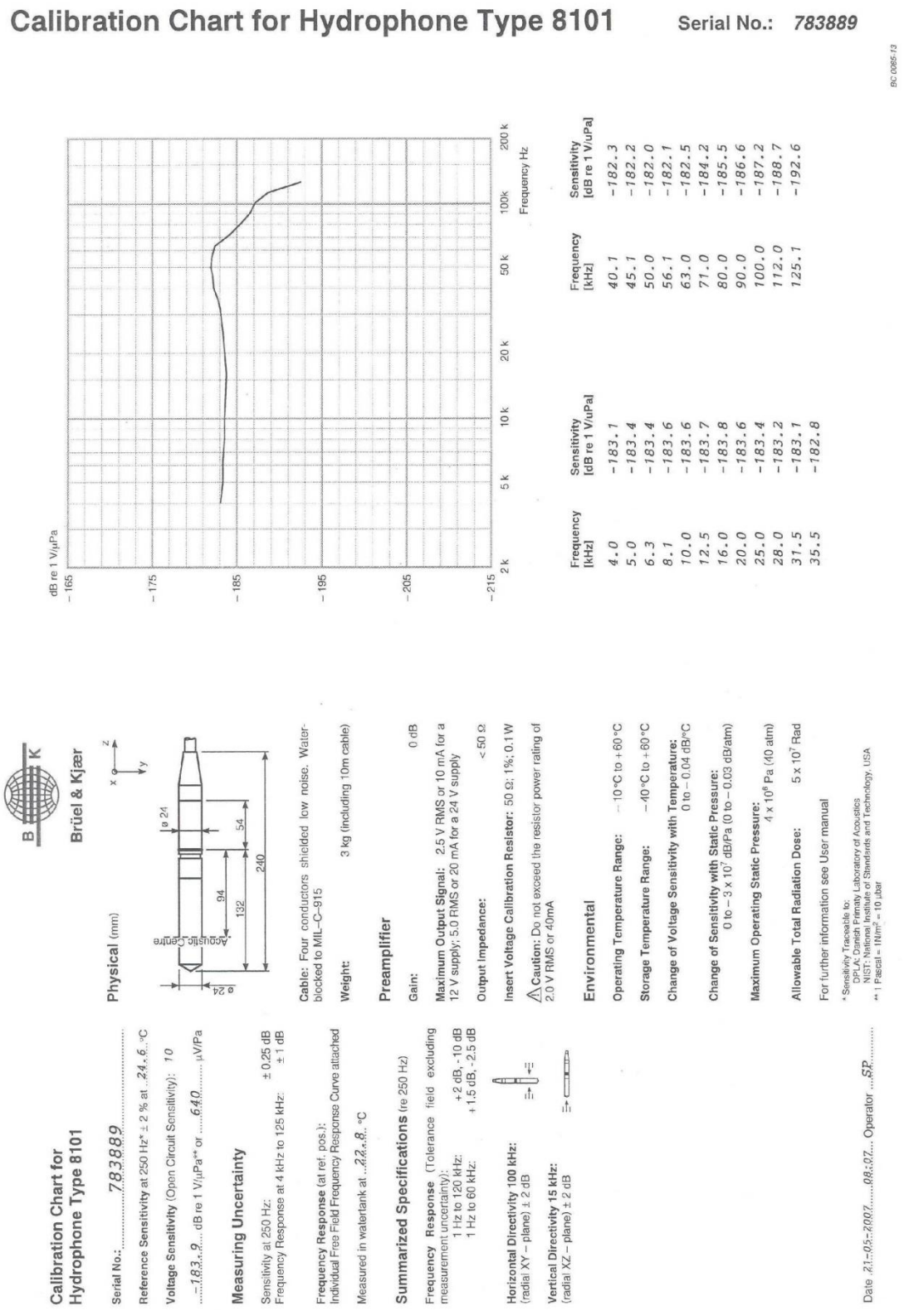
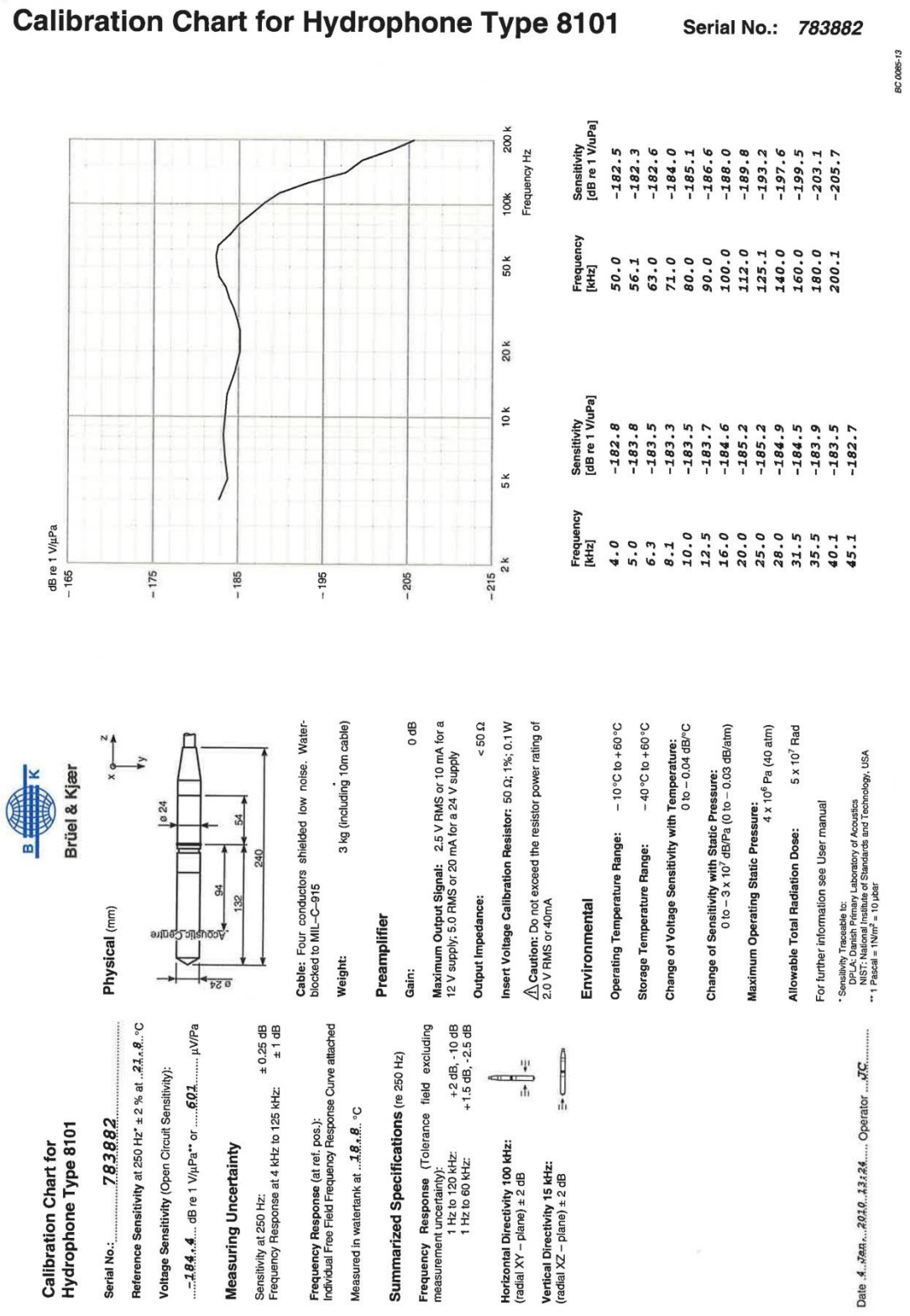


Fig. 8. Calibration chart for the reference hydrophone used in 2012



Calibration Chart for Hydrophone Type 8101

Serial No.: *783882*

BC 0085-13

Brüel & Kjær

Calibration Chart for Hydrophone Type 8101

Serial No.: *783882*

Reference Sensitivity at 250 Hz* ± 2 % at *21.8* °C

Voltage Sensitivity (Open Circuit Sensitivity):

-184.4 dB re 1 V/µPa** or *6.01* µV/Pa

Measuring Uncertainty

Sensitivity at 250 Hz: ± 0.25 dB
Frequency Response at 4 kHz to 125 kHz: ± 1 dB

Frequency Response (at ref. pos.):
Individual Free Field Frequency Response Curve attached

Measured in watertank at *18.8* °C

Summarized Specifications (re 250 Hz)

Frequency Response (Tolerance field excluding measurement uncertainty):
1 Hz to 120 kHz: +2 dB, -10 dB
1 Hz to 60 kHz: +1.5 dB, -2.5 dB

Horizontal Directivity 100 kHz: (radial XY – plane) ± 2 dB

Vertical Directivity 15 kHz: (radial XZ – plane) ± 2 dB

Date *4. Jan. 2010* Operator *VC*

Physical (mm)

Cable: Four conductors shielded low noise. Water-blocked to MIL–C–915

Weight: 3 kg (including 10m cable)

Preamplifier

Gain: 0 dB

Maximum Output Signal: 2.5 V RMS or 10 mA for a 12 V supply; 5.0 RMS or 20 mA for a 24 V supply

Output Impedance: < 50 Ω

Insert Voltage Calibration Resistor: 50 Ω; 1%; 0.1 W

⚠ **Caution:** Do not exceed the resistor power rating of 2.0 V RMS or 40mA

Environmental

Operating Temperature Range: – 10 °C to + 60 °C

Storage Temperature Range: – 40 °C to + 60 °C

Change of Voltage Sensitivity with Temperature: 0 to – 0.04 dB/°C

Change of Sensitivity with Static Pressure: 0 to – 3 x 10^7 dB/Pa (0 to – 0.03 dB/atm)

Maximum Operating Static Pressure: 4 x 10^6 Pa (40 atm)

Allowable Total Radiation Dose: 5 x 10^7 Rad

For further information see User manual

* Sensitivity Traceable to:
DPLA: Danish Primary Laboratory of Acoustics
NIST: National Institute of Standards and Technology, USA

** 1 Pascal = 1N/m² = 10 µbar

dB re 1 V/µPa
– 165
– 175
– 185
– 196
– 205
– 215
2 k
5 k
10 k
20 k
50 k
100k
200 k
Frequency Hz

Frequency [kHz]	Sensitivity [dB re 1 V/uPa]	Frequency [kHz]	Sensitivity [dB re 1 V/uPa]
4.0	*-182.8*	*50.0*	*-182.5*
5.0	*-183.8*	*56.1*	*-182.3*
6.3	*-183.5*	*63.0*	*-182.6*
8.1	*-183.3*	*71.0*	*-184.0*
10.0	*-183.5*	*80.0*	*-185.1*
12.5	*-183.7*	*90.0*	*-186.6*
16.0	*-184.6*	*100.0*	*-188.0*
20.0	*-185.2*	*112.0*	*-189.8*
25.0	*-185.2*	*125.1*	*-193.2*
28.0	*-184.9*	*140.0*	*-197.6*
31.5	*-184.5*	*160.0*	*-199.5*
35.5	*-183.9*	*180.0*	*-203.1*
40.1	*-183.5*	*200.1*	*-205.7*
45.1	*-182.7*		

Table 9. Calibration chart data for the reference hydrophone used in 2013

Hydrophone sensitivity for Reson TC4032-1
Date: 2007-05-31
S/N: 5003146

Frequency (kHz)	Sensitivity (dB re 1 V/uPa)
5	-171.3
6	-171.5
7	-171.8
8	-172.1
9	-172.5
10	-173.2
11	-173.8
12	-174.3
13	-174.2
14	-173.8
15	-173.6
16	-173.5
17	-173.4
18	-173.2
19	-173.0
20	-172.6
22	-172.2
24	-172.2
26	-172.1
28	-172.3
30	-172.3
32	-171.9
34	-171.4
36	-171.1
38	-170.7
40	-170.4
42	-170.2
44	-170.1
46	-170
48	-169.7
50	-169.4

Frequency (kHz)	Sensitivity (dB re 1 V/uPa)
52	-169
54	-168.6
56	-168.7
58	-168.9
60	-169.3
62	-169.6
64	-169.9
66	-170
68	-170
70	-170.2
72	-170.4
74	-170.7
76	-171
78	-171.5
80	-171.9
84	-172.8
88	-173.6
92	-174.4
96	-175.2
100	-175.6