THE COMPANY OF
Biologists

RESEARCH ARTICLE

Equal latency contours and auditory weighting functions for the harbour porpoise (*Phocoena phocoena*)

Paul J. Wensveen[1,*], Léonie A. E. Huijser[2], Lean Hoek[2] and Ronald A. Kastelein[2]

ABSTRACT

Loudness perception by human infants and animals can be studied under the assumption that sounds of equal loudness elicit equal reaction times (RTs). Simple RTs of a harbour porpoise to narrowband frequency-modulated signals were measured using a behavioural method and an RT sensor based on infrared light. Equal latency contours, which connect equal RTs across frequencies, for reference values of 150−200 ms (10 ms intervals) 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 higher the signal level was above the hearing threshold of the harbour porpoise, the quicker the animal responded to the stimulus (median RT 98−522 ms). Equal latency contours roughly paralleled the hearing threshold at relatively low sensation levels (higher RTs). The difference in shape between the hearing threshold and the equal latency contours was more pronounced at higher levels (lower RTs); a flattening of the contours occurred for frequencies below 63 kHz. Relationships of the equal latency contour levels with the hearing threshold were used to create smoothed functions assumed to be representative of equal loudness contours. Auditory weighting functions were derived from these smoothed functions that may be used to predict perceived levels and correlated noise effects in the harbour porpoise, at least until actual equal loudness contours become available.

KEY WORDS: Audiogram, Effects of noise, Frequency weighting, Odontocete, Response latency

INTRODUCTION

Concern about the effects of anthropogenic noise on marine mammals has led to attempts to establish acoustic safety criteria for underwater noise (Southall et al., 2007). Noise criteria often contain some form of frequency-selective weighting according to the perception of the target species, so that single thresholds apply to many sounds irrespective of their frequency spectra. Until recently, such weighted thresholds for marine mammals were 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 obtained using auditory weighting functions derived from equal loudness contours (e.g. A- and C-weightings) (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). Recently, Finneran and Schlundt (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 reviews, see 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; Pfingst et al., 1975b), house finch (*Carpodacus mexicanus*) (Dooling et al., 1978) and domestic cat (*Felis catus*) (May et al., 2009); near-threshold contours have been obtained for the harbour seal (*Phoca vitulina*) (Kastelein et al., 2011). 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. If auditory weighting based on RT improves predictions of noise effects in marine mammals, this method may be a relatively time-efficient alternative to auditory weighting based on direct loudness estimates.

In this study, underwater equal latency contours were measured in a harbour porpoise, *Phocoena phocoena* (Linnaeus 1758), responding behaviourally to narrowband frequency-modulated (FM)

[1]Sea Mammal Research Unit, University of St Andrews, St Andrews, Fife KY16 8LB, UK. [2]Sea Mammal Research Company (SEAMARCO), Julianalaan 46, 3843 CC Harderwijk, The Netherlands.

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

The Journal of Experimental Biology

List of symbols and abbreviations

DAQ	data acquisition
FM	frequency modulation
HF	high frequency
I	sound intensity
I_0	sound intensity of the hearing threshold
L_{ht}	SPL of the hearing threshold
L_{lat}	SPL of the smoothed equal latency contour
L_{loud}	SPL of the equal loudness contour
LED	light-emitting diode
LF	low frequency
MF	mid-frequency
RMSE	root mean square error
RT	reaction time
SnL	sensation level
SPL	sound pressure level
TTS	temporary threshold shift
W	weighting level

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.

RESULTS

A total of 5144 trials were conducted in 167 experimental sessions, resulting in 3822 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. 1 as functions of both sensation level (SnL) (*sensu* Ellison et al., 2012) 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 log–log slopes towards the middle frequencies (Table 1, Fig. 1). The goodness of fit values were satisfactory: the coefficient of determination (r^2) ranged from 0.90 to 0.99 and the root mean square error (RMSE) ranged from 2.7 to 6.4 ms (Table 1).

Six equal latency contours (I−VI) were constructed from the auditory RT functions. Equal latency contours V and VI (corresponding to 190 and 200 ms) roughly followed the shape of the hearing threshold, including the notch in the audiogram at 63 kHz (Fig. 2). 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 log–log slopes of the auditory RT functions (Table 1, parameter α).

The six equal latency contours were converted into hypothetical equal loudness contours (Fig. 3) and auditory weighting functions (Fig. 4); the parameter estimates for Eqns 3 and 4 are provided in Table 2. 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 10 to 16 dB per octave, depending on the equal latency contour it was based upon.

DISCUSSION
Evaluation of the RT data
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 1.5–5 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) 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 150–200 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 lower 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 and colleagues (Wagner et al., 2004). For SnLs between 0 and 40 dB, plotting the median RTs for all frequencies in a similar graph as Fig. 1 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 (Chocholle, 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
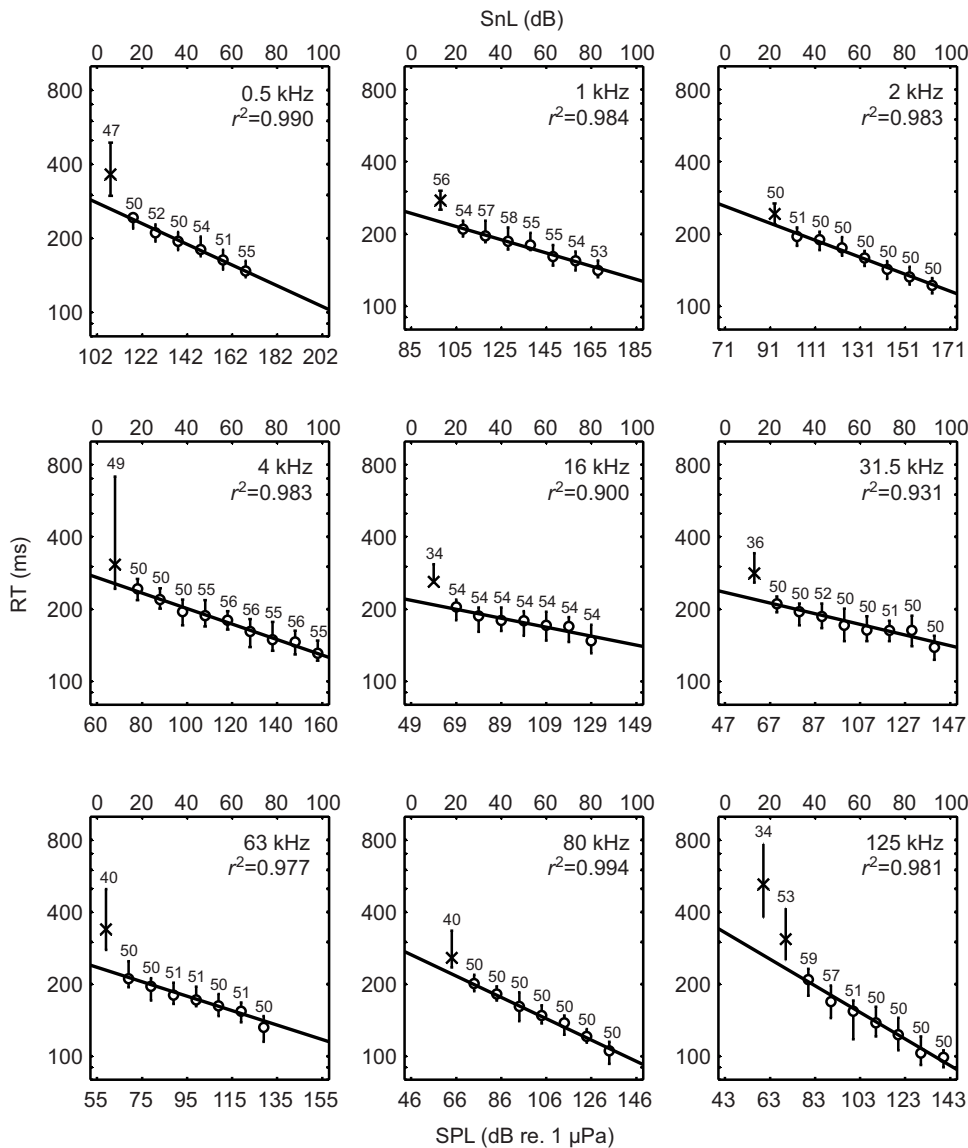
Fig. 1. Relationship between sensation level (SnL; top axes), sound pressure level (SPL; bottom axes) and reaction time (RT) of the harbour porpoise. Nine narrowband frequency-modulated (FM) tonal signals with centre frequencies of 0.5–125 kHz were tested. The auditory RT functions (black lines; Eqn 1) result from fitting a power law (Eqn 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. Note the logarithmic scale on the ordinate.

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

Table 1. Best-fit parameters for the auditory RT functions (Eqn 1)

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

Shown in the table are 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 1), and the goodness-of-fit parameters r^2 and root mean square error (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 sound pressure level (SPL) and sensation level (SnL) in Fig. 1.
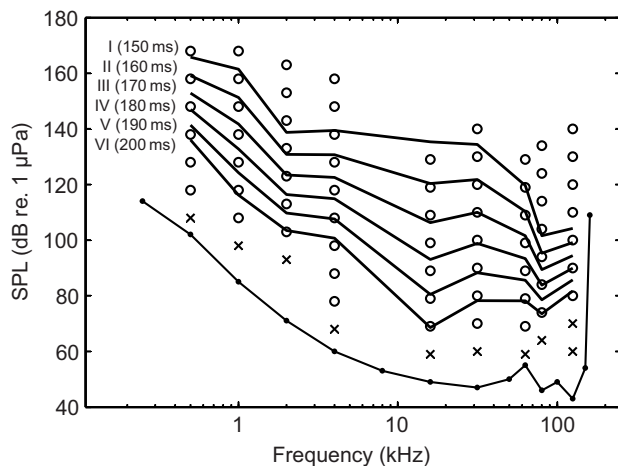
Fig. 2. Equal latency contours I−VI of the harbour porpoise with the corresponding reference RT values. The contour SPLs were derived from the auditory RT functions in Fig. 1 by matching the six reference RTs across frequencies. Circles and crosses indicate the test levels for which median RTs were included 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).

changes in the audiogram over a 3–4 year period (Kastelein et al., 2013). In this more recent study, the porpoise was not allowed to echolocate during research trials.

Relationship between RT 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., 2011; 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 (Sanders, 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 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 et al., 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
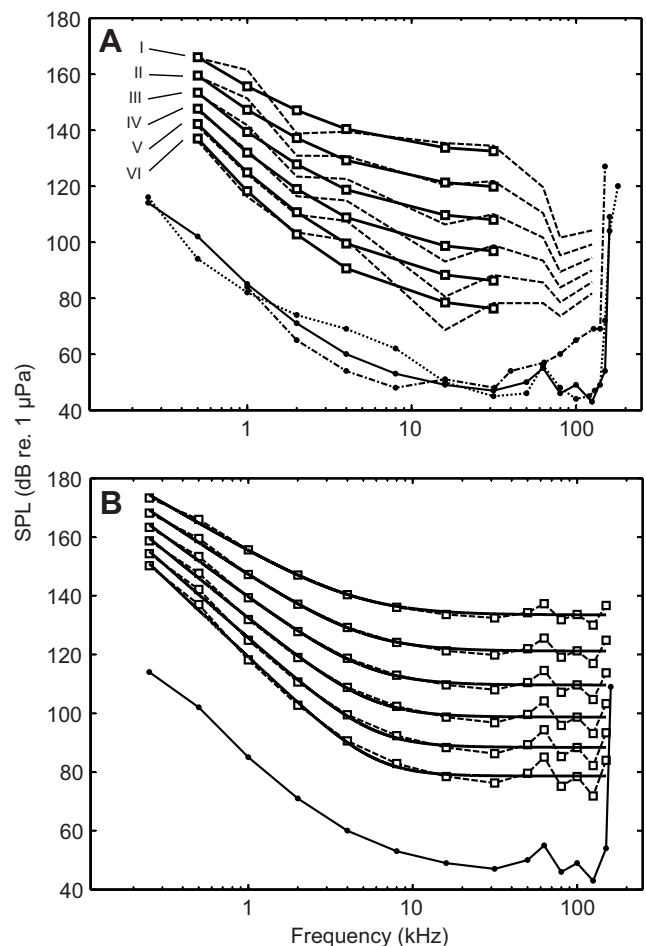

Fig. 3. Conversion from equal latency contours to equal loudness contours. (A) The shape of the harbour porpoise's audiogram was used as a template to create smoothed versions of the equal latency contours (solid lines with squares) from the original contours (dashed lines; same as in Fig. 2). The audiogram of the subject [solid line with dots (Kastelein et al., 2010)] and the audiograms of two other harbour porpoises [dashed–dotted line (Andersen, 1970) and 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 2; see Materials and methods, 'Derivation of the auditory weighting functions' for parameter values), which resulted in six hypothetical equal loudness contours (dashed lines with squares). A closed-form model (Eqn 3) was fitted (solid lines) to these extended smoothed contours. Only the audiogram of the subject is shown here (solid line with dots).

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 and colleagues (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
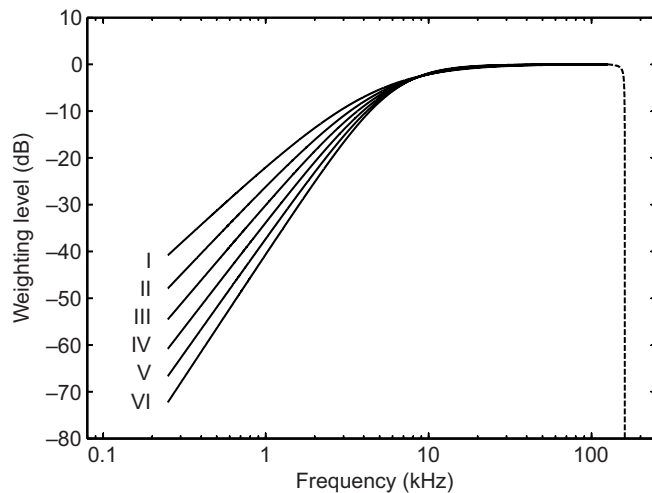
Fig. 4. Six auditory weighting functions for the harbour porpoise. The weighting functions, which should reflect the frequency response of the porpoise's hearing system, are associated with the SnL of the received signal (I being high and VI being low). The extension of the functions (dashed line) is to emphasize that the effective hearing range of the porpoise ends abruptly at 160 kHz.

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. 2), 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 elsewhere (Green, 1975; May et al., 2009; Kastelein et al., 2011) were also lower 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.

Ecological significance and recommendations

The six auditory weighting functions (Fig. 4) are assumed to represent relative loudness perception in the porpoise. The experimental method used in this study is relatively fast compared with 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) is associated with the loudest sounds. TTS is generally induced by loud sounds; therefore, a function relating TTS onset levels to frequency is expected to be similar in shape to the inverse of the flattest weighting curve. In contrast, 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). This may suggest that TTS onset levels are not always perceived as equally loud across frequencies. When data on TTS onset in harbour porpoises become available for multiple frequencies, the method of Finneran and Schlundt (Finneran and Schlundt, 2013) may be used to determine which of the six auditory weighting functions is the most appropriate for predicting TTS onset in this species. Similarly, behavioural response threshold SPLs may be used to determine which of the weighting functions is the most appropriate for predicting specific behavioural effects, in cases where these effects are highly correlated with loudness.

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 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.

The weighting functions in this study may be used to predict behavioural response thresholds independent of the frequency of the

Table 2. Parameter estimates with 95% CIs for the hypothetical equal loudness contours (Eqn 3; Fig. 3B) and the auditory weighting functions (Eqn 4; Fig. 4)

Contour	K_1 (dB re. 1 μPa) Best-fit value	95% CI	K_2 (dB) Best-fit value	95% CI	a (Hz) Best-fit value	95% CI	x Best-fit value	95% CI	Roll-off rate (dB octave^{-1})
I	133.48	3.32	0.03	3.32	4563	3322	1.62	0.58	9.7
II	121.17	3.80	0.02	3.80	5004	4180	1.84	0.65	11.1
III	109.62	4.27	0.01	4.27	5318	4393	2.05	0.71	12.3
IV	98.71	4.73	0.01	4.73	5550	4550	2.26	0.78	13.6
V	88.41	5.17	0.00	5.17	5722	4667	2.45	0.84	14.7
VI	78.62	5.60	0.00	5.60	5856	4758	2.64	0.90	15.9

Data are shown for parameters K_1, K_2, a and x (see Materials and methods, 'Derivation of the auditory weighting functions'). 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.
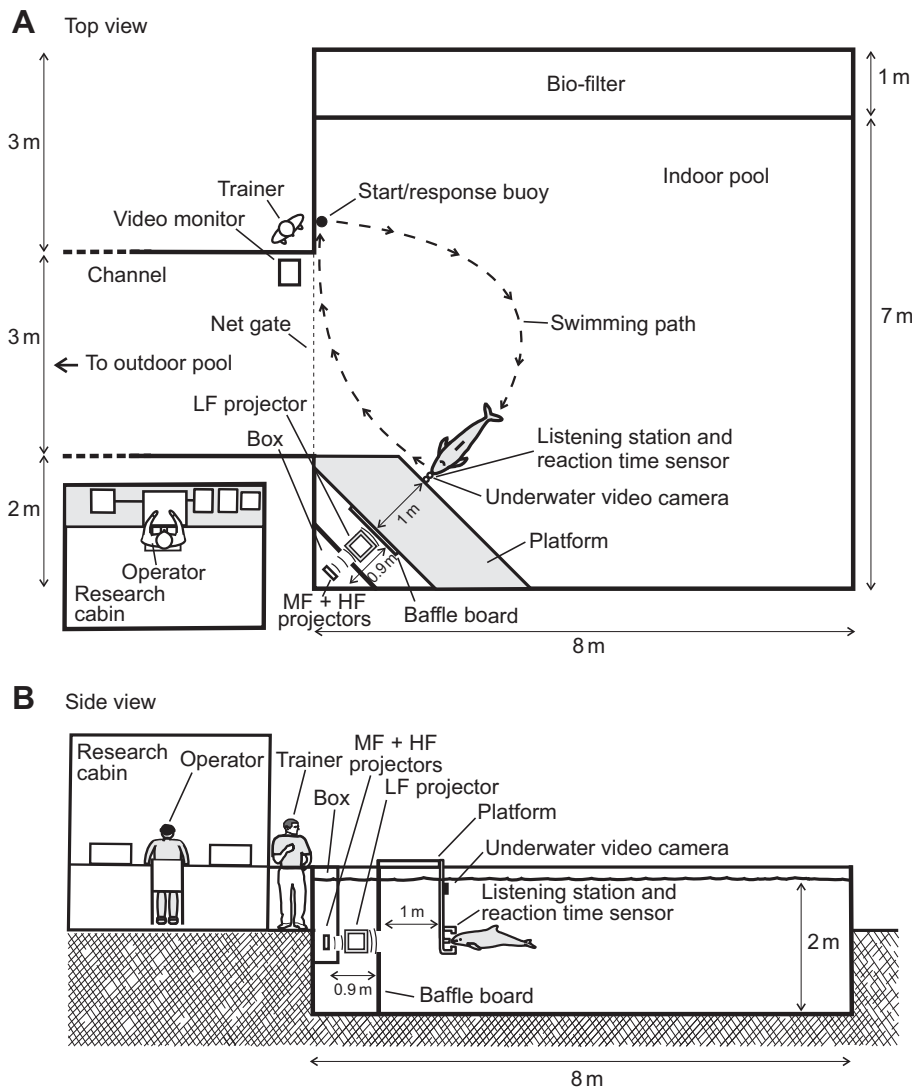
Fig. 5. The indoor test pool. The pool is shown (A) from above and (B) from the side. The harbour porpoise, trainer and signal operator are shown in their positions at the beginning of a research trial. Signals were transmitted with a low frequency (LF; 0.5−4 kHz), mid-frequency (MF; 16−63 kHz) or high frequency (HF; 80−125 kHz) projector.

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

MATERIALS AND METHODS
Study animal
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 a recent study on TTS (Kastelein et al., 2012). 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 (8×7 m, 2 m deep; Fig. 5) 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 min 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. 5). 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).

The research protocol was approved by the University of St Andrews' School of Biology Ethics Committee. Animal training and data collection 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.

Table 3. The 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

Contour	Reference RT (ms)	RMSE (dB)				
		At ≤16 kHz	At ≤31.5 kHz	At ≤63 kHz	At ≤80 kHz	At ≤125 kHz
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

Test frequencies included in the data sets for the similarity analyses (see 'Derivation of the auditory weighting functions' for more details) are given above each RMSE column. The smoothed contour was the transformed hearing threshold (Eqn 2) that was the most similar to the unsmoothed contour. The highest similarity between the two curves, indicated by the smallest RMSE (in italics), was generally obtained with the 0.5−31.5 kHz data set.

Test 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 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 SnL. SnL is defined here as the number of dB above the subject's 50% detection hearing threshold for 900 ms tonal signals, which was measured 2–3 years ago (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.

Sound production and monitoring

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. 6. 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. 6). 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 ≥48 dB, but in other sessions, only 4 kHz signals of SnLs ≤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) (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. 5), 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. 5). 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 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 and Kjær (B&K) 2635, Nærum, Denmark] and amplified loudspeaker, or a modified ultrasound detector (Batbox III, Steyning, UK).

Reaction time measurements

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 Fig. 5B). 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.
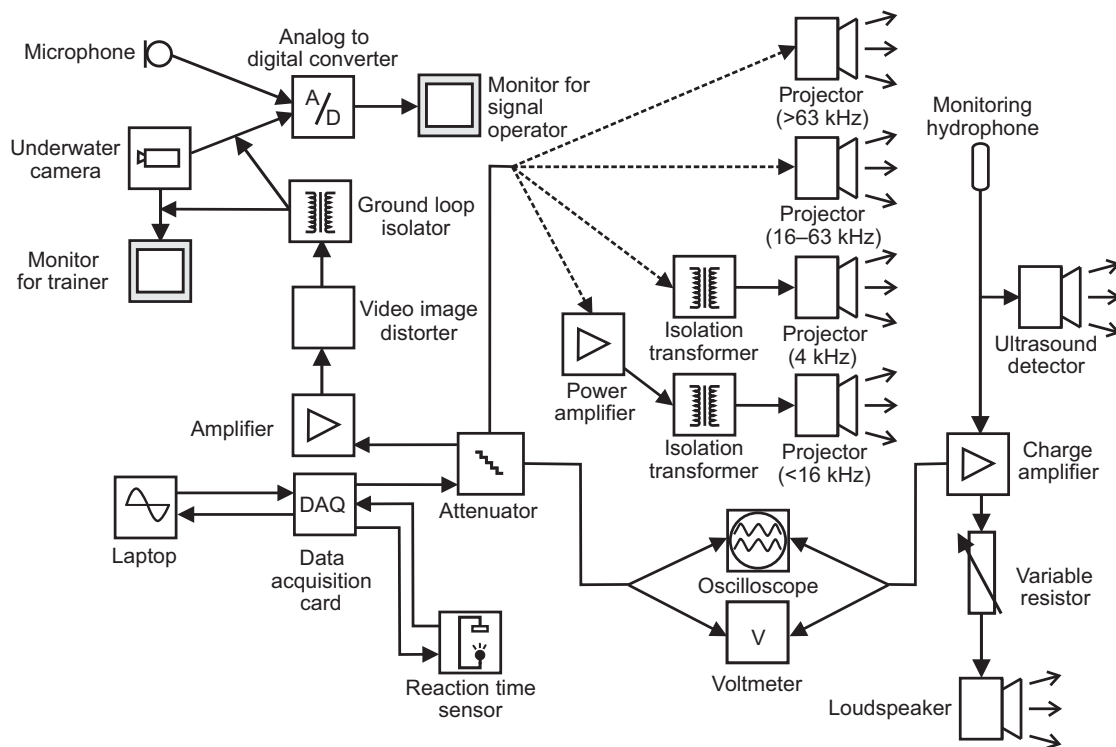
Fig. 6. Equipment set-up. Block diagram of the equipment used to generate the sound stimuli, monitor the sounds and the harbour porpoise under water, and measure the porpoise's RTs.

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. 5). 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.

Acoustic calibration
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. 7A) and positioned his rostrum against it, so that his anterior–posterior axis was aligned with the acoustic beam axis of the projector (Fig. 7B). 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. 7C) upon detecting either the test stimulus or the trainer's whistle by returning to the start/response buoy (Fig. 7D), 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. 7C), 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
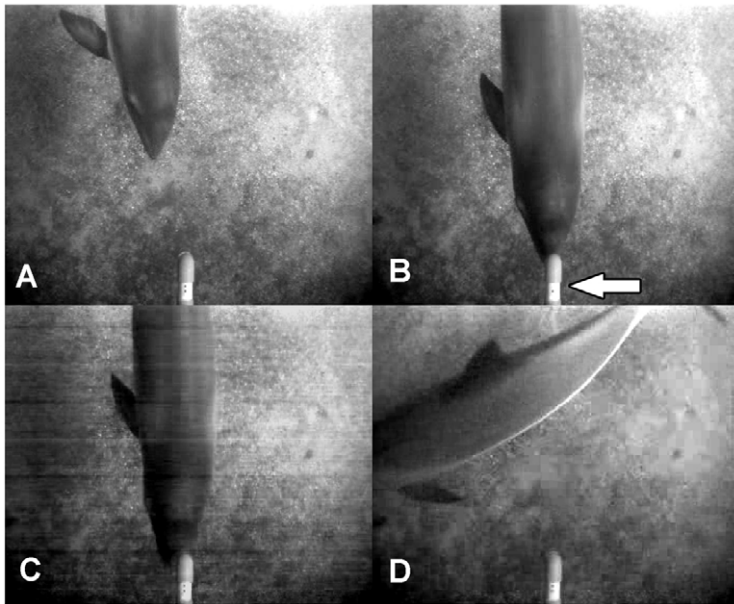
Fig. 7. Still frame sequence of a research trial. (A) The harbour porpoise approaches the listening station. (B) The porpoise positions its rostrum against the listening station, thereby breaking the infrared light beam in the reaction time sensor, and illuminating the LED in the top sensor bracket above 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.

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 5 days a week in 2011 (sessions started at 09:00 h, 11:00 h and 14:00 h), and one extra session was performed daily in 2012 (starting at 16:00 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.

Analysis of the reaction times

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:

$$RT = \beta(I/I_0)^{-\alpha} \; , \qquad (1)$$

where I/I_0 is the ratio of the intensity of the test stimulus (I) to the intensity of a stimulus at threshold (I_0) (calculated using $I/I_0=10^{SnL/10}$), 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 (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. 1). 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.

Derivation of the auditory weighting functions

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 (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}):

$$L_{lat}(f) = \gamma L_{ht}(f) + \delta \; , \qquad (2)$$

where f is a vector of test frequencies, and γ and δ are a scaling and translation parameter, respectively. The 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 3). However, the similarity between the unsmoothed and smoothed versions of contours I−IV increased significantly after the exclusion of high frequency data (Table 3), 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 for parameter γ were: 0.610, 0.721, 0.825, 0.924, 1.016 and 1.104 dB/dB and for parameter δ: 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 1) (Kastelein et al., 2010) in the calculated contour/threshold relationships (Eqn 2). This resulted in a family of six hypothetical equal loudness contours. A closed-form model was fitted to each of these contours:

$$L_{\text{loud}}(f) = K_1 - 20\log_{10}\left[\frac{b^x f^x}{(a^x + f^x)(b^x + f^x)}\right], \quad (3)$$

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 et al., 2000; Southall et al., 2007) also used by Finneran and Schlundt (Finneran and Schlundt, 2011) is a special case of Eqn 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 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 a very high number (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 3 were normalized to 0 dB at the most sensitive frequency and inverted to obtain the auditory weighting functions, which then take the form:

$$W(f) = K_2 + 20\log_{10}\left[\frac{b^x f^x}{(a^x + f^x)(b^x + f^x)}\right], \quad (4)$$

where $W(f)$ is the weighting level in dB, a, b and x are as in Eqn 3, and K_2 is a vertical offset that results from the normalization of the contours ($K_2 = \min\{L_{\text{loud}}\} - K_1$).

Author contributions
P.J.W. and R.A.K. designed the experimental study and gained funding. P.J.W., L.A.E.H. and L.H. collected and processed the reaction time data. P.J.W. analysed the data and prepared the manuscript with advice on results and interpretation from all authors.

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 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 dissertation, Texas A&M University, USA.
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. *Annee 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. *Journal de Psychologie.* **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. In *Proceedings of the 155th Meeting of the Acoustical Society of America*, pp. 117-122.
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. 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 dolphins (*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., 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., 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 0.315 and 150 kHz in random Gaussian white noise. *J. Acoust. Soc. Am.* **126**, 1588-1597.
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 0.25 and 160 kHz. *J. Acoust. Soc. Am.* **128**, 3211-3222.

Kastelein, R. A., Wensveen, P. J., Terhune, J. M. and de Jong, C. A. F. (2011). 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., Gransier, R., Hoek, L. and Olthuis, J. (2012). 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. (2013). 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.

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.

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.

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. *Aquatic Mammals* **38**, 362-401.

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**, 1368-1379.

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

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.

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. *Annee Psychol.* **22**, 58-142.

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

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

Ridgway, S. H., Carder, D. A., Kamolnick, 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, C. R., Jr, Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E. et al. (2007). Marine mammal noise exposure criteria: initial scientific recommendations. *Aquatic Mammals* **33**, 411-414.

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.

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 *Proceedings of the Undersea Defence Technology Conference and Exhibition.* Amsterdam: Nexus Business Media.

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.