



# Relationship between sperm whale (*Physeter macrocephalus*) click structure and size derived from videocamera images of a depredating whale (sperm whale prey acquisition)

Delphine Mathias, Aaron Thode, Jan Straley, Kendall Folkert

## ► To cite this version:

Delphine Mathias, Aaron Thode, Jan Straley, Kendall Folkert. Relationship between sperm whale (*Physeter macrocephalus*) click structure and size derived from videocamera images of a depredating whale (sperm whale prey acquisition). *Journal of the Acoustical Society of America*, 2009, pp.3444-3453. 10.1121/1.3097758 . hal-00904060

**HAL Id: hal-00904060**

**<https://hal.science/hal-00904060>**

Submitted on 13 Nov 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Relationship between sperm whale (*Physeter macrocephalus*) click structure and size derived from videocamera images of a depredating whale (sperm whale prey acquisition)

Delphine Mathias<sup>a)</sup> and Aaron Thode

Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Jan Straley

University of Alaska Southeast, Sitka, Alaska 99835

Kendall Folkert

Alaska Longline Fishermen's Association, P.O. Box 6497, Sitka, Alaska 99835

(Received 12 August 2008; revised 10 February 2009; accepted 18 February 2009)

Sperm whales have learned to depredate black cod (*Anoplopoma fimbria*) from longline deployments in the Gulf of Alaska. On May 31, 2006, simultaneous acoustic and visual recordings were made of a depredation attempt by a sperm whale at 108 m depth. Because the whale was oriented perpendicularly to the camera as it contacted the longline at a known distance from the camera, the distance from the nose to the hinge of the jaw could be estimated. Allometric relationships obtained from whaling data and skeleton measurements could then be used to estimate both the spermaceti organ length and total length of the animal. An acoustic estimate of animal length was obtained by measuring the inter-pulse interval (IPI) of clicks detected from the animal and using empirical formulas to convert this interval into a length estimate. Two distinct IPIs were extracted from the clicks, one yielding a length estimate that matches the visually-derived length to within experimental error. However, acoustic estimates of spermaceti organ size, derived from standard sound production theories, are inconsistent with the visual estimates, and the derived size of the junk is smaller than that of the spermaceti organ, in contradiction with known anatomical relationships. © 2009 Acoustical Society of America. [DOI: 10.1121/1.3097758]

PACS number(s): 43.80.Ka, 43.30.Sf [WWA]

Pages: 3444–3453

## I. INTRODUCTION

The question of whether an organism's anatomical dimensions can be inferred from features of its acoustic signal has a long history in bioacoustics. The bulk of this research has focused on inferring the length of the mammalian vocal tract via formant analysis or other spectral measures.<sup>1–5</sup>

Attempts to derive a cetacean's body size from its acoustic signal structure face even greater challenges than terrestrial studies, due to the difficulty of obtaining independent and accurate measurements of animal dimensions in the wild. The most detailed theory linking a cetacean sound to an individual's anatomical dimensions concerns sperm whales (*Physeter macrocephalus*), which produce a 25–30 ms transient sound called a click. Researchers have divided sperm whale sounds into a variety of categories, based on the timing between subsequent clicks in a sound sequence. This timing pattern, called the inter-click interval (ICI), has been used to classify clicks into usual clicks, slow clicks, creaks, and codas.<sup>6</sup> Clicks are believed to serve a variety of functions, including echolocation and communication, depending on the click pattern. Usual clicks typically have an ICI of 0.5–1.0 s, while creaks display a shorter ICI between 0.2 s and 0.5 s.<sup>7,8</sup> ICI values within a given creak also tend to

decrease with time, and orientation measurements from tagged animals have demonstrated that sudden changes in orientation are associated with creak sounds.<sup>8,9</sup> These observations, among others, suggest that creaks are used as an echolocation signal.<sup>6,7,10–14</sup>

A click displays internal structure in the form of several local maxima, or “pulses,” with the time interval between pulses within a click labeled the “inter-pulse interval” (IPI). Note that this quantity is different from the ICI defined earlier.

A 1972 paper by Norris and Harvey<sup>10</sup> hypothesized that a click is initially generated at a pair of fatty tissues called the *museau de singe*, at the front of the animal's nose.<sup>42</sup> Under this hypothesis, most acoustic energy escapes the nose directly, while the remaining fraction propagates backwards through the spermaceti organ, gets reflected forward by the frontal air sac, and finally escapes into the water either via the anterior end of the spermaceti organ or the rostrum (Fig. 1).

Modifications to this model have been made by Mohl<sup>15–17</sup> and supported by Zimmer *et al.*<sup>18,19</sup> Under this “bent-horn” interpretation, the initial omnidirectional pulse  $P_0$  transmitted directly into the water is actually only a small portion of the sound generated at the *museau de singe*, whereas the bulk of the energy propagates backwards through the spermaceti organ and reflects off the frontal air sac. The largest component of this reflection is transmitted

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: delphine.mathias@gmail.com

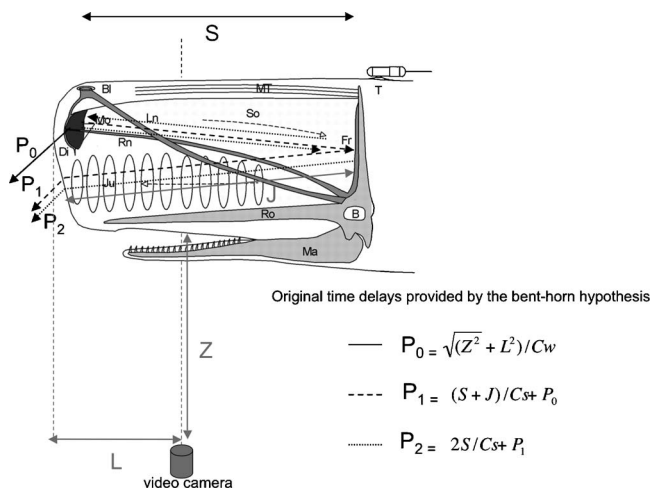


FIG. 1. Bent-horn model of sound production (adapted from Fig. 1 of Ref. 43) and associated formulas for acoustic path lengths. Anatomical labels: B: brain, BI: blow hole, Di: distal air sac, Fr: frontal air sac, Ju: junk, Ln: left naris, Ma: mandible, Mo: monkey/phonics lips or museau de singe, MT: muscle/tendon layer, Rn: right naris, Ro: rostrum, So: spermaceti organ. Propagation path variables: S: length of the spermaceti organ, J: length of the junk, Z: distance between the videocamera and the jaw of the animal, L: lateral distance between the nose and the videocamera,  $C_w$ : sound speed in the water=1500 m/s,  $C_s$ : sound speed in the spermaceti organ=1370 m/s.

into the water via the junk, creating a highly directional main pulse  $P_1$ . Recent measurements<sup>19</sup> also indicate that a portion of the energy reflected from the frontal air sac (just over the skull) escapes directly into the water, creating a  $P_{1/2}$  pulse that can be detected between the  $P_0$  and  $P_1$  pulse in recordings made off the longitudinal axis of the animal. Finally, under the revised theory, a portion of the  $P_1$  pulse energy propagates back into the junk and/or spermaceti organ, reflects off the frontal sac, and passes a second time through the junk into the water, creating a  $P_2$  pulse. These propagation paths are illustrated in Fig. 1.

According to both the “Norris–Harvey” and bent-horn theories, the nominal IPI, or IPI between  $P_1$  and  $P_2$ ,<sup>20</sup> is proportional to the two-way acoustic travel time between the *museau de singe* and frontal sac, provided that the measurements are made either directly in front of or behind the animal, such that the aspect-dependent  $P_{1/2}$  pulse merges with the  $P_0$  or  $P_1$  pulse. In 2005, Zimmer *et al.*<sup>19</sup> showed that measurements made off the longitudinal axis yield off-axis “distortion,” in the sense that the appearance of a  $P_{1/2}$  pulse obfuscates the interpretation of the timing measurements.

A natural consequence of these hypotheses is that the nominal IPI could be used to estimate the length of the spermaceti organ or junk. Allometric relationships between nose size and body length<sup>21</sup> then suggest that the nominal IPI should be correlated with body length. Indeed, functional regressions between the nominal IPI and body length have been published.<sup>22–24</sup> These polynomials were derived using analyzed IPIs from whales whose length had been independently measured while they had surfaced. However, no independent means of estimating the animals’ spermaceti organ size were available.

Various automated methods have been explored for estimating the nominal IPI, but cepstral analysis<sup>25</sup> has been used in several published papers,<sup>20,23,26</sup> under the assumption

that a click can be modeled as a convolution of a “source” and a “reflection” function. Teloni *et al.*<sup>20</sup> showed that aspect-dependent features of the IPI estimates could be effectively removed by averaging a large number of cepstra derived from clicks recorded during an animal’s foraging dive, during which the animal presumably presents a wide variety of orientations with respect to the recording hydrophone, thus permitting aspect-dependent features of the clicks to be averaged down and the nominal IPI to be enhanced.

This paper uses an unusual dataset to directly compare visual estimates of the size of a sperm whale’s head with acoustic estimates of its total length, spermaceti organ length, and junk length. Under normal foraging conditions sperm whales typically dive to depths greater than 300 m,<sup>8,9,27–29</sup> making it impossible to acquire video recordings of prey acquisition attempts. However, sperm whales have learned how to depredate fishing gear, particularly demersal longline operations, in a number of locations around the globe since 2002,<sup>30–34</sup> including Norway, Greenland, eastern Canada (Labrador and Newfoundland), Chile, and the Falkland Islands. It is the largest marine mammal known to depredate on human fishing activities, and these activities have received increasing coverage in the scientific literature.<sup>31–39</sup>

In 2003 the Southeast Alaska Sperm Whale Avoidance Project (SEASWAP) was established by scientists, managers, and fishermen to characterize the severity of sperm whale depredation activity on black cod (*Anoplopoma fimbria*) off Sitka, Alaska. Passive acoustic measurements collected during SEASWAP discovered that the animals occasionally dove below the fishing vessels at depths less than 100 m, depths presumably shallow enough to permit visual observations of this activity. In 2006, videocameras were deployed for the first time to capture this behavior.

Section II describes the video and acoustic recording equipment, discusses how the camera was deployed from a fishing vessel during an active longline haul, and then outlines the procedures used to derive visual and acoustic estimates of the size of an animal captured on-camera. Section III describes a specific depredation encounter recorded on May 31, 2006, during which the whale touches the fishing line at a known distance from an underwater videocamera lowered to 108 m depth. The visual and acoustic estimates of the animal’s size are compared, and the timing between a particular set of pulses within a click yields an acoustic length estimate that corresponds well with the visually-derived length estimate. The results confirm earlier studies that the IPI can be correlated with total body length, but finds inconsistencies in the visual and acoustic estimates of spermaceti organ size.

## II. PROCEDURE

### A. Video and audio recording equipment

Visual data were collected by a Sony HVR-1AU videocamera, housed in a Gates Underwater Products HC1/A1U housing using a WP-25 lens port (80° field of view, 63.5 mm diameter lens). The data were recorded onto Sony PHDVM-

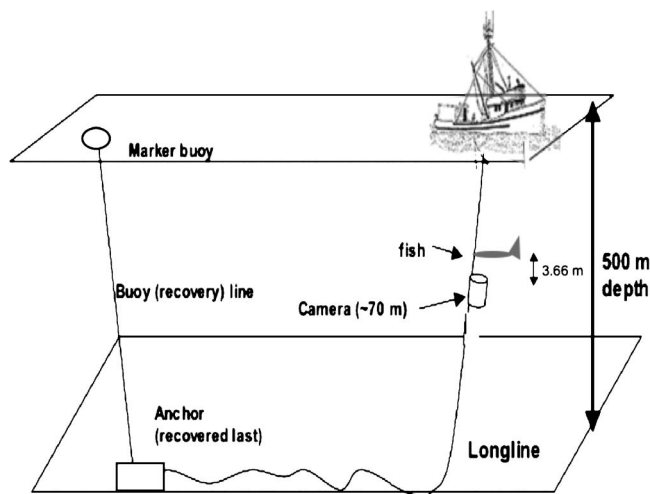


FIG. 2. Schematic of camera deployment from fishing vessel.

63DM DigitalMaster tapes in DVCAM format, recording over 60 min of uncompressed audio and video per tape. The housing and camera port were depth-rated to 152 m (500 ft). The widest angle field of view was used for the recordings (minimum zoom).

The Gates housing also contained a Kobitone PN 252-LM049 Electret Condenser Microphone, recorded to an audio channel on the videotape simultaneously with the video-stream. The official sensitivity range of the microphone was 20–12 kHz, with  $-162 \text{ dB re } 1 \text{ V}/\mu\text{Pa}$  sensitivity. The data were sampled at 48 kHz.

## B. Deployment procedure

All deployments took place from the F/V Cobra, a 59 ft fishing vessel mastered by one of the authors (Folkert), who also designed most of the camera deployment system (Fig. 2). A standard demersal longline is comprised of 200 m long lengths of 1/4 in. groundline called “skates,” and at every meter along a skate a baited hook is attached by a small line called a “ganion,” typically about 20 cm long. Under normal operations the longline is typically deployed by dropping a surface buoy over the side, deploying sufficient line to account for the water depth and desired scope of the buoy, and then deploying a 30 kg anchor overboard. As the fishing vessel moves over the desired site, the baited longline pays out the back of the vessel. After a typical set has been cast (about 4–6 km long) a second anchor is deployed overboard, attached to a second surface buoy. After a 3–12 h “soak,” the upstream surface buoy is recovered, and the anchor line is passed over a “roller” mounted on the side of the vessel and through a hydraulically-operated pot-puller to haul the rope up and recover the fish. It is at this point that whales like to depredate the line, as fish are accessible throughout the water column.

The deployment technique was designed to avoid substantial changes in fishing procedures, which would have run into regulatory issues. Thus “blank” skates, marked at 50 m intervals, were inserted between every two baited skates, permitting a normal deployment and haul. During the recovery, a blank skate would be hauled until the start of the next

hooked skate began to emerge from the water. If whales were present, the camera assembly was activated, sealed, and attached to the start of the hooked skate, with the lens port oriented so that it would be facing the ocean surface. The camera was attached to the line so that two to six fish (already present on the line) would be visible above the camera. The distance of each fish from the camera and the length of each fish was recorded and noted. The assembly would then be lowered down into the water until tape marks on the line indicated the desired deployment depth had been reached by the camera (Fig. 2). The true deployment depth could be measured precisely by attaching a commercial dive computer to the line just underneath the camera. Thus the entire procedure minimized alterations to standard fishing procedures.

## C. Audio data analysis

Acoustic data from the camera were extracted from the video and stored as a 48 kHz 16-bit WAV file. A peak detector was used to locate echolocation clicks. For every local temporal maximum detected, a 30 ms click sample was extracted and saved, centered around the arrival time of the signal peak. This length of time window was chosen to avoid contamination by preceding or subsequent clicks.

The time difference between subsequent clicks was used to estimate the ICI. By plotting ICI as a function of time, click detections that shared the same ICI “trajectory,” and thus presumably belonging to the same whale, could be selected for further analysis, removing clicks generated by other whales in the vicinity.

Once the click samples from the camera sequence were isolated, the IPI was then extracted from each time sample using two different methods: an incoherent peak detection method, applied to the Hilbert transform of the signal, and via the signal cepstrum,<sup>26</sup> which presumes that the click can be expressed as the convolution between a scattering/reflection function and an impulsive “excitation” function.

To obtain the first estimate of the click envelope, a Hilbert transform was applied to each click time series, and then the transforms were stacked on top of each other to permit a view of the click evolution. The time origin was defined as the arrival time of the most intense pulse within the click, which will be labeled as pulse B. The envelopes were averaged over every 0.6 s of the recording to produce estimates evenly sampled in time. To determine pulse intervals, the local maxima in the [5–10] ms interval before and after pulse B were flagged, using methods nearly identical to the original peak detection method used to detect clicks in the data. These intra-pulse maxima were then labeled A and C. (The conventional notation of  $P_0$ ,  $P_1$ , etc., used by the bent-horn hypothesis is avoided here in order to avoid a particular interpretation of the pulse structure). The precision in the pulse estimate was typically 0.05 ms.

The second measurement method used cepstral analysis. Heuristically speaking, the cepstrum measures any periodic oscillation of the signal spectrum with respect to frequency, which can be interpreted in the time domain as the time delay between an original signal and its reflection. Thus for the case of a sperm whale click, the Norris–Harvey hypoth-



esis and later modifications predict that the cepstrum, which is measured in units of time, should display a peak at the time corresponding to the two-way travel time between the *museau de singe* and the frontal sac.<sup>23</sup>

After band-pass filtering between 2 and 18 kHz, cepstral estimates were made over the entire 30 ms window, the -15 to 0 ms window, and the 0–15 ms window, relative to the arrival time of pulse B. This was done in order to estimate separate IPI values for the A-B and B-C pulse intervals. Caution must be exercised when band-pass filtering a time series with a finite impulse response (FIR) filter before computing the signal cepstrum, since the application of the filter produces ripples in the output spectrum, which then produces an artificial peak in the cepstrum. It was found that by specifying a wide filter stopband of 1000 Hz, cepstral artifacts were restricted to 0–3 ms band in the cepstral output, outside the region of interest. Cepstral segments were then averaged over every 0.6 s of the recording (as with the Hilbert transform plots) to produce estimates evenly sampled in time.

Finally, the whale body length can be estimated from the IPI using the Gordon polynomial relating sperm whale body length and extracted IPI:<sup>23</sup>

$$\text{Total length (m)} = 4.833 + 1.453\text{IPI(s)} - 0.001\text{IPI}^2(\text{s}). \quad (1)$$

Rhineland and Dawson also derived a relationship between sperm whale size and IPI using photogrammetric length estimates, and measurements of the IPI for 12 individuals:<sup>24</sup>

$$\text{Total length (m)} = 17.120 - 2.189\text{IPI(s)} + 0.251\text{IPI}^2(\text{s}). \quad (2)$$

Gordon used measurements from 11 individuals in the tropical Indian Ocean, while the Rhineland and Dawson dataset contained 66 whales from Kaikoura (New Zealand). Gordon's regression contained only one individual longer than 12 m, while all individuals from Rhineland and Dawson's dataset were larger than 12 m, possibly explaining the large difference between the polynomial coefficients between Eqs. (2) and (3). Thus, Rhineland and Dawson's polynomial would be expected to be more suited for sperm whales living in the Gulf of Alaska, which are generally males greater than 12 m.

#### D. Video analysis

The dataset discussed here involves video images of a whale contacting a fishing line at a known distance of 3.66 m (12 ft) from the videocamera, with its body oriented roughly perpendicular to the camera. The underexposed features of the whale were enhanced by performing a power law gray scale transformation ( $\gamma=0.5$ ). The animal was judged to be perpendicular to the camera plane when the teeth on the left side of the lower jaw lined up in such a way as to block the view of corresponding teeth in the right side of the jaw.

On shore the camera was attached to a rope and oriented so that when it was lowered to 3.66 m depth in a pool, the camera pointed toward the surface. A trellis of known size

(1.22 m on a side) was slid on the surface across the camera's field of view to permit a conversion of pixel separation to physical distance.

The camera was further calibrated for image distortion using a printed checkboard pattern of 1.2 m width, 1 m height, with 10 cm squares. The board was placed 3 m from the camera, and was moved in many directions in order to get as many angular views as possible. Then 20 images were extracted and standard camera calibration procedures were used to extract the system's focal length, principal point, and image distortion.<sup>40</sup> The net result is a distortion model that maps the radial and tangential distortion of every pixel in the image. It was found that pixels that lie within half the image width or height from the image center suffered less than 1% distortion, while pixels on the image borders experienced 4% distortion. The second estimate of distortion was made by placing the camera at 3.66 m water depth and measuring at the size variation of the trellis when moved across the camera's field of vision. In this case, it was found that the size of the trellis suffered less than 3% distortion when placed on the image border.

The combined scaling and distortion calibrations permitted three physical measurements to be extracted from the image: the distance between the tip of the snout to the point where the jaw hinges to the skull, or "gape angle" (SG), the distance between the tip of the nose to the start of the upper jaw (SJ), and the mean spacing between the animal's teeth (TS). Ideally, the distance between the blowhole and the center of the eye should have been measured (BE), since this distance is a good estimate of the spermaceti organ length,<sup>41</sup> and thus the IPI estimate could be directly related to spermaceti organ size. However, due to the low level of ambient illumination the eye cannot be located in the silhouette.

Allometric relationships published in Ref. 21 relate the SG and the SJ to total body length, thus permitting an estimate of the size of the skull and the length of the whale, even if the entire animal is not visible in the video.<sup>42</sup> Unfortunately, the SG measurement is not precise, because the actual location of the gape angle lies under the animal's skin, resulting in a possible error of 30%–50% in animal size. Thus the third measurement, TS, was made.

From the video it is possible to measure the distance between teeth along portions of the jaw, including the locations of the first four teeth near the tip of the jaw, and the rearmost nine teeth. To determine whether an allometric relationship exists between total body length and mean tooth spacing, tooth measurements were obtained from three male sperm whale skeletons with original total body lengths 18.3, 11.9, and 14 m. Two skeletons were from beached animals stored by the Natural History Museum of Los Angeles County in Southern California, and thus originated from the Pacific Ocean, and the third was from the Nantucket Historical Association Whaling Museum in Nantucket, Massachusetts. The tooth separation varies with distance along the jaw, so the mean value of the spacing was computed using the teeth that correspond to the teeth visible in the image. The ratio between the total measured length and mean tooth spacing (MLTL) was then computed. A corresponding MLTL can be estimated from the image, using the TS measured from

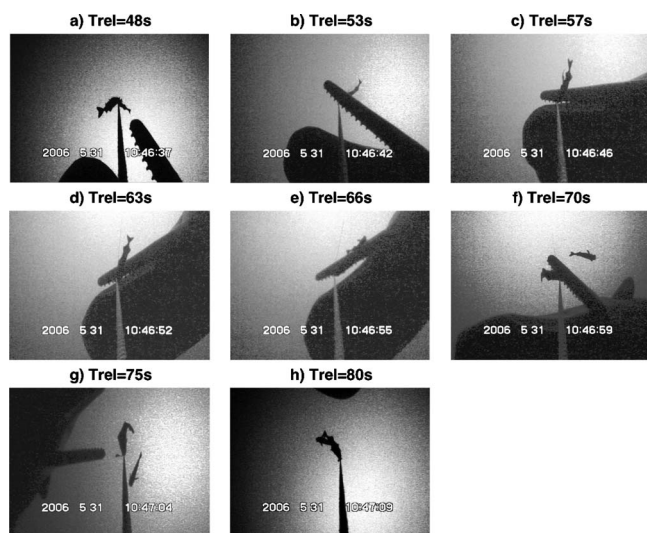


FIG. 3. Snapshots from the May 31 sperm whale encounter video, with the relative time in seconds counted from 10:45:59. Figures are labeled (a) through (h), starting from upper left and moving left-to-right across columns. (f) was used for the visual length estimate.

the image and the body length derived from the SJ and TS measurements, to determine whether the video MLTL lies within the ranges of the skeletons' MLTL.

The BE dimension, and thus the spermaceti organ size, was interpolated from the SJ, SG, and TS measurements, using the data of Nishiwaki *et al.* Anatomically speaking, the SG dimension is always larger than the BE dimension, which means an upper bound can be placed on the size of the spermaceti organ from the image analysis.

### III. RESULTS

#### A. General description of the May 31st 2006 encounter

The first videotaped encounter of a depredating sperm whale took place on May 23 2006; however, the whale never contacted the rope, precluding a visual estimate of its body size. A second encounter was recorded on May 31, 2006, during one of the last attempts to obtain such a recording. The deployment depth of the camera was 108 m, and two black cod were present at distances of 3.66 and 5.49 m (12 and 18 ft) from the camera, respectively. The camera was activated at 10:25 local time, and by 10:45 had been lowered to the target depth. Significant sperm whale acoustic activity was recorded on the videotape from the moment the camera entered the water.

In this and all following sections, the relative timing of events will be expressed as seconds elapsed from 10:45:49, the time at which the acoustic activity from the animal initiated off camera. At 48 s [Fig. 3(a)], the jaw of a sperm whale appears and contacts the longline at an estimated distance of 1.8 m from the camera. The animal then slides along the longline until its jaw lies adjacent to the fish nearest the camera (3.66 m) at 53 s [Fig. 3(b)]. The animal then seems to completely close its jaw around the longline at that point by 57 s [Fig. 3(c)], deflecting the rope a considerable amount. At 63 s [Fig. 3(d)], the animal performs a slight

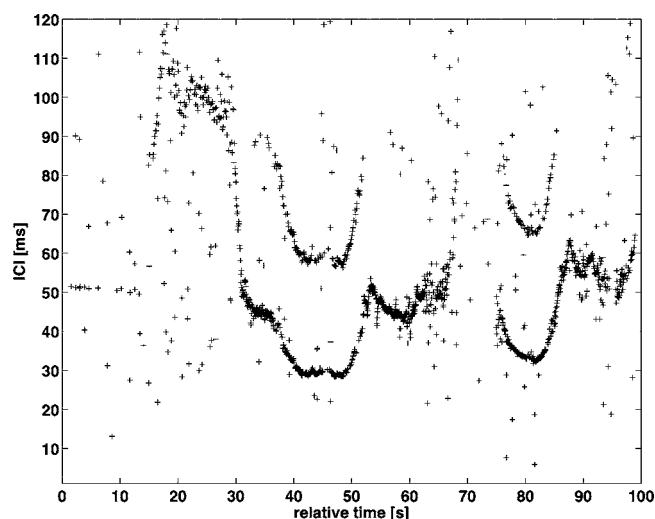


FIG. 4. ICI for the May 31 camera sequence, with relative time referenced in seconds from 10:45:49.

barrel roll, and the fish attached a distance of 5.49 m from the camera breaks off the longline and floats away. The fish immediately adjacent to the jaw does not detach. As the animal opens its jaw, it stops producing creak clicks and works itself free of the line between 66, 70, and 75 s [Figs. 3(e)–3(g)]. Figure 3(f) is the key image used in subsequent analysis; it is the moment when the whale's head is judged perpendicular to the camera. Once free, the animal and floating fish are seen on-camera until 80 s, during which the whale seems to be orienting its head toward the loose fish. Unfortunately whatever happens next occurs after both the whale and fish float out of view.

#### B. Acoustic analysis

The click detection procedure in Sec. II C generated 1723 click samples from the May 31 sequence over roughly 100 s. Figure 4 displays the ICI between individual clicks, expressed in the relative time scale. The ICI was estimated simply by measuring the time interval between successive detections in the detection function. Note that the time between 30 s and 85 s corresponds to ICI values of under 50 ms, traditionally considered characteristic of “creak” activity (e.g., Refs. 6 and 9), and thus most of the sequence analyzed here involves high SNR creak clicks.

As an aside, note that even at times when the whale is biting down on the line, the animal still produces creak clicks, and these clicks indicate a minimum ICI of 30 ms, or about 33/s, considerably lower than the maximum click rate of 90.9 click/s in other accounts.<sup>7</sup> The timing of these pulses is consistent with a two-way travel time from a target 22 m away, although the whale is clearly interested in targets just a couple of meters away. Thus Fig. 4 suggests the whale may be reaching a physiological limit in its click production rate. The lack of a visible ICI between 68 and 75 s corresponds to the time when the whale is working its jaw free of the line, and the ICI sequence reappears a fraction of a second after the jaw snaps free from the line. This correspondence between acoustic and visual events provides confidence that the clicks analyzed here were produced by the whale viewed on

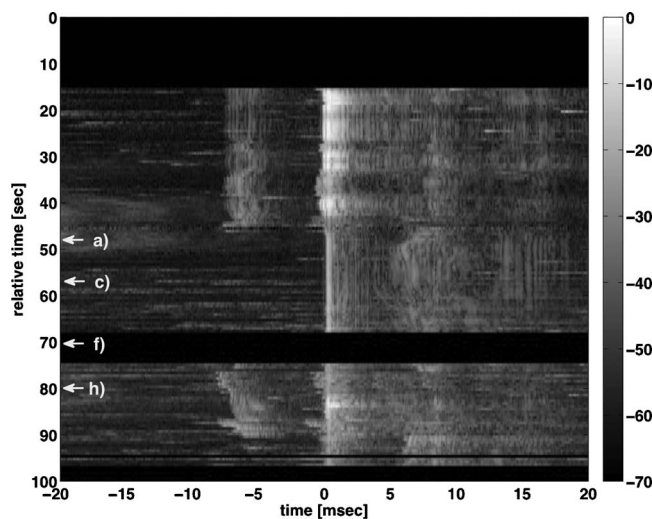


FIG. 5. Stacked click structure of the May 31 sequence, created by filtering each click sample between 2 and 18 kHz, applying the Hilbert transform, and averaging transforms in 0.6 s bins. Time on the y-axis is expressed relative to 10:45:49.

the camera. Figure 4 was thus used to window clicks to those that fit this ICI pattern, and thus the on-camera whale, leaving 1178 clicks.

Figure 5 displays the resulting stacked plot of the absolute values of the Hilbert transform of the filtered click time series for the May 31st sequence, following the procedure in Sec. II C, with key events labeled using overlying letters that correspond to the images from Fig. 3. As previously discussed, each transform has been time-shifted so that the signal maximum (B pulse) aligns with the time origin.

One aspect of Fig. 5 that attracts instant attention is the sudden disappearance of either the A or B pulse from the train of clicks between 45 and 75 s, which corresponds closely to the times that the depredating whale appears on-camera. In the rest of this discussion, click samples from within this time range are dubbed “ambiguous” clicks, because of the ambiguity in associating the missing pulse with either the  $P_0$  or  $P_1$  pulse of the bent-horn hypothesis. All other clicks outside this time range are dubbed “standard” clicks. Figure 6 shows representative spectrograms of ambiguous and standard clicks. From this point on, the body length analysis uses only standard click samples.

As stated in Sec. II C, two cepstra were computed from each standard click: one only containing the A and B pulses, and one only containing the B and C pulses. The individual click cepstra then needed to be averaged to produce a clear IPI peak, as was also found by Teloni *et al.*<sup>20</sup> Cepstral averages over 5 and 10 s time intervals were examined, but reliable cepstral estimates were only obtained by averaging all standard cepstra.

Figure 7(a) displays the stacked cepstra generated from the entire click, Fig. 7(b) displays the average cepstrum using only the [A-B] portion of the standard clicks, and Fig. 7(c) displays the average cepstrum using only the [B-C] portion of the standard clicks. In Fig. 7(b), a weak local maximum at 6.3 msec is discernable, while in Fig. 7(c) a strong local maximum at 8.15 ms is clearly visible.

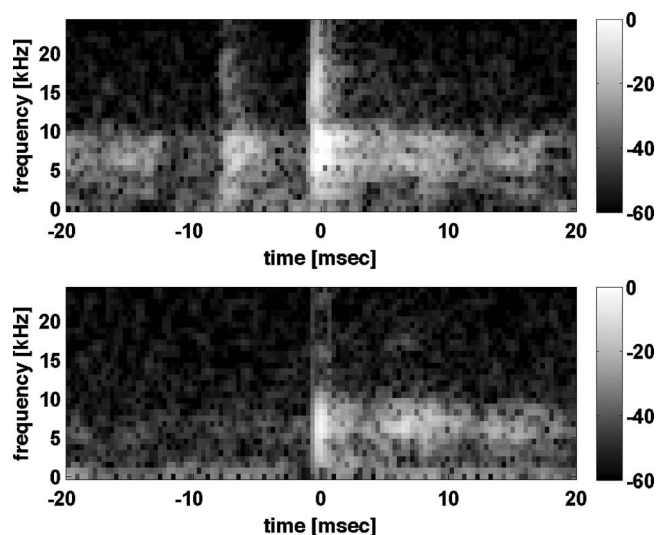


FIG. 6. Representative spectrogram of (a) standard click (42 s in Fig. 5) and (b) ambiguous click (55 s in Fig. 5).

Table I summarizes all IPI estimates extracted from the standard clicks, derived using both Hilbert transform and cepstral methods. Bootstrap methods were used to estimate the variance of the cepstral estimates, by randomly selecting 850 individual cepstra from the appropriate set of click types, averaging, and measuring the peak.

### C. Body length estimation using visual and acoustic methods

In Fig. 3(f) the entire whale’s head is visible in the image, oriented perpendicularly with respect to the camera, from the tip of the nose to the point where the lower jaw enters the head. The point where the jaw touches the line is 3.66 m from the camera.

Figure 8 displays the superposition of Fig. 3(f) and an image from the pool calibration, from which the distance between the snout tip to the best estimate of gape angle (SG) is  $3.8(\pm 0.1)$  m, and the distance between the snout tip to the lower jaw tip (SJ) is  $1.07(\pm 0.04)$  m. The mean tooth separation is  $0.100(\pm 0.02)$  m with an uncertainty of 0.005 m on the measurements. These estimate uncertainties arise from both uncertainties in the exact location of the gape in the image, as well as distortion effects. To check how these measurements were affected by the specific image chosen, the measurements were repeated on two other video images extracted 0.5 s before and after Fig. 3(f), when the whale is not exactly perpendicular to the camera. The resulting total length estimates lie within 0.12 m of the results obtained from Fig. 3(f).

Figure 9, derived using whaling data from Ref. 21, relates SG, SJ, and BE measurements as a function of total whale length. As mentioned above, the BE measurement provides an estimate of spermaceti organ length, and the SG measurement provides an upper bound on this organ’s length. A  $3.8(\pm 0.1)$  m SG length translates into a total body length estimate of  $15.2(\pm 0.3)$  m, while a  $1.07(\pm 0.04)$  m SJ length translates into a total body length estimate of  $15.25(\pm 0.06)$  m, a good agreement. The resulting BE length



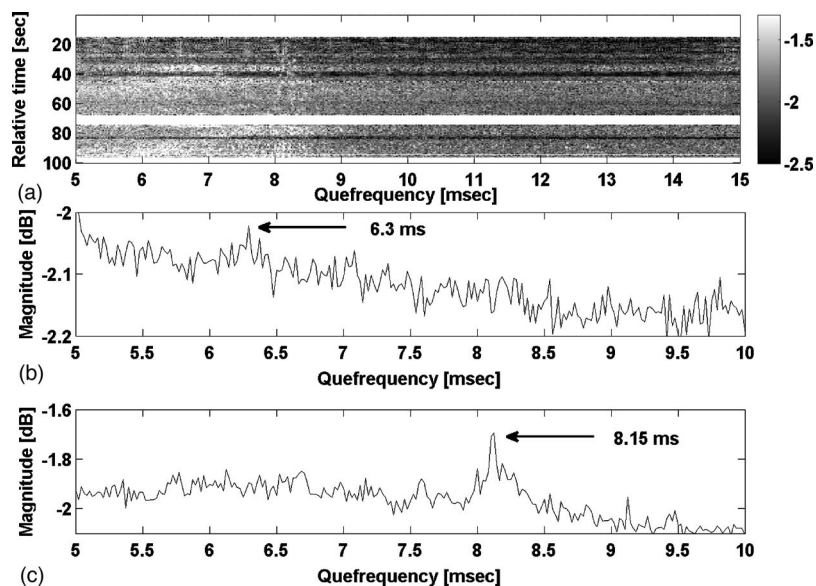


FIG. 7. (a) Stacked cepstra of May 31st sequence, computed over the set of standard clicks (0–48 s in Fig. 5); (b) averaged cepstrum of the [A-B] portion of the standard click samples; (c) averaged cepstrum of the [B-C] portion of the standard click samples.

estimate corresponds to a  $3.4(\pm 0.1)$  m spermaceti organ. As mentioned in Sec. II D, concerns about the accuracy of the SG measurement prompted additional measurements of the ratio between the derived body length and mean tooth separation. This dimensionless ratio was  $152(\pm 38)$ , and the ratios derived from the skeletons described in Sec. II D are  $179(\pm 27)$ ,  $160(\pm 23)$ , and  $168(\pm 16)$ . Thus the SJ and SG measurements yield virtually identical results for total length, and the teeth measurements produce the same length estimate to within experimental error, alleviating concerns that the SG measurement might yield length estimates that are 40%–50% below the actual value.

Both the Gordon and the Rhineland and Dawson poly-

nomials, Eqs. (1) and (2), are combined with the two IPI estimates (Table I) to yield four acoustic estimates of the body length. The acoustic estimate obtained by applying the Rhineland and Dawson formula to the B-C measurements fits the visual estimate to within a meter, as would be expected from previous literature. An acoustic estimate of the spermaceti organ size can be derived, assuming that the B-C interval corresponds to the  $P_1$ – $P_2$  interval of the bent-horn hypothesis, which in turn represents the two way travel time within the spermaceti organ. An estimate of the junk size can then be derived by assuming the A-B interval corresponds to the  $P_0$ – $P_1$  interval, which yields the combined acoustic path length of the spermaceti organ and junk.

TABLE I. Body length estimation for May 31st sequence: comparison between visual and acoustic data.

	Acoustic measurements					
	IPI (ms)	Body length Eq. (1) (m)	Body length Eq. (2) (m)	Anatomical dimension (m)	Body length Fig. 9 (m)	Anatomical description
B-C component, Hilbert	7.6 ( $\pm 0.9$ )	15.9 ( $\pm 1.1$ )	15.1 ( $\pm 1.0$ )	5.23 ( $\pm 0.6$ )		Spermaceti organ
B-C component, cepstral	8.1 ( $\pm 0.2$ )	16.5 ( $\pm 0.3$ )	15.8 ( $\pm 0.3$ )	5.55 ( $\pm 0.3$ )		Spermaceti organ
A-B component, Hilbert	6.4 ( $\pm 0.8$ )	14.1 ( $\pm 1.0$ )	13.4 ( $\pm 1.1$ )	8.8 ( $\pm 1.0$ )		Spermaceti + junk
A-B component, cepstral	6.3 ( $\pm 0.5$ )	13.9 ( $\pm 0.5$ )	13.3 ( $\pm 0.6$ )	8.6 ( $\pm 0.6$ )		Spermaceti + junk
Video measurements						
SG estimate				3.8 ( $\pm 0.1$ )	15.2 ( $\pm 0.3$ )	Tip of snout to angle of gape
SJ estimate				1.07 ( $\pm 0.04$ )	15.25 ( $\pm 0.06$ )	Tip of snout to tip of lower jaw
BE estimate				3.4 ( $\pm 0.1$ )		Blowhole to center of eye
Teeth estimate				0.100 ( $\pm 0.02$ )	15.2 ( $\pm 0.3$ )	Mean teeth separation





FIG. 8. Superimposition of the videocamera image [Fig. 3(f)] with a pool calibration image.

Table I compares the derived body lengths for both the visual and acoustic methods (the latter using standard clicks only). A 8.1 ms B-C interval and 6.3 ms A-B interval yields respective spermaceti and junk lengths of 5.55 and 3.1 m.

## IV. DISCUSSION

### A. Relationship between IPI and anatomical structure

Table I shows that both the cepstral and Hilbert IPI estimation procedures produce similar results, in that they find that the A-B IPI is at least 1 ms shorter than the B-C interval. When this IPI difference is applied to either Eq. (1) or (2), the different IPI values yield a 2 m difference in whale length. The body length computed from the B-C time interval, using the Rhinelander and Dawson polynomial [Eq. (2)] matches the body length estimated from the video (15.2 m) to within the 0.6 m experimental uncertainty, which indicates that the B-C “nominal” IPI ( $P_1$ - $P_2$  interval) must have dominated the IPI measurements used to derive the original poly-

nomial fit. Equation (2) would be expected to provide a better fit than Eq. (1), as all individuals from Rhinelander and Dawson dataset were larger than 12 m, whereas Gordon’s regression contained only one individual longer than 12 m. Thus to this point, the observations are consistent with the bent-horn hypothesis.

However, the videocamera data raise some questions about the physical interpretation of the IPI in the animal. Under the bent-horn hypothesis, the  $P_1$ - $P_2$  interval should represent the two-way travel time of a sound pulse inside the spermaceti organ; thus the acoustic measurements indicate a spermaceti organ on the order of 5.5 m [assuming a propagation speed of 1370 m/s (Refs. 13 and 26)]. However, the visual estimate of the spermaceti organ length, using the BE measurement from the image, is only 3.4 m (Table I and Fig. 9), which is only 60% of the acoustic estimate. One simple explanation for the discrepancy is that the BE measurement (from the blowhole to the center of the eye) is a biased estimate of spermaceti organ length. Whether that bias would be so large to yield almost a factor of 2 error is unknown.

Another anatomical puzzle is that the 3.1 m length of the junk derived by Sec. III C is shorter than the 5.55 m spermaceti organ length, which is in contradiction with known anatomical relationships.<sup>41</sup> Were the museau de singe physically further from the camera than the acoustic exit point from the junk, this discrepancy might be explained; however, the physical orientation of the whale relative to the camera indicates that the relative distance between the two expected acoustic exit points should not be a large effect.

In summary, the combined video/acoustic measurements suggest that the IPI of the sperm whale can be related to the total length of the animal, as has been demonstrated empirically elsewhere many times. However, the video observations are inconsistent with standard interpretations of the propagation paths through the animal’s head; namely, that the  $P_1$ - $P_2$  interval is a direct measure of the size of the spermaceti organ.

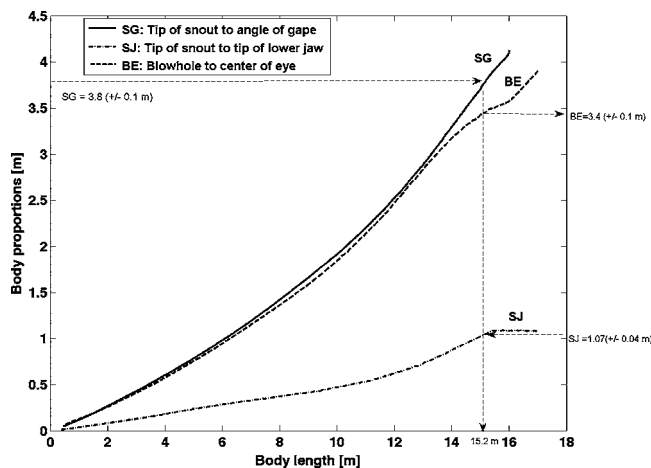


FIG. 9. Allometric relationships between body proportions of male sperm whales caught in the North Pacific (modified from Ref. 21). Solid line: SG, distance from tip of snout to angle of gape; dot-dash line: SJ, distance from tip of snout to tip of lower jaw; dashed line: BE, distance from blowhole to center of eye. Both SG and SJ were measured directly from the image. The BE measurement is a proxy for spermaceti organ length.

### B. Unresolved questions

The dataset shown in Fig. 5 shows some additional puzzling features that are worth mentioning briefly. The first is the lack of a so-called  $P_{1/2}$  pulse, observed in detailed measurements elsewhere.<sup>19</sup> The orientation of the animal captured by the videocamera would suggest a clear time separation between the  $P_0$ ,  $P_{1/2}$ , and  $P_1$  pulses, yet our results indicate no trace of this additional propagation path.

Even more puzzling is the temporal structure of the so-called ambiguous clicks, visible between the 48 and 80 s time window in Fig. 5. Out of the two longline encounters captured by videocamera, this type of click structure only appears in the May 31 encounter. The ambiguous clicks are intriguing because they are produced when the whale is biting the line next to a fish, and have a highly variable ICI (Fig. 4). Why would a whale continue to creak or “buzz” while biting the line, when the presumed targets of interest are off-axis of the presumed sonar beam and actually behind the museau de singe?

Furthermore, the ambiguous clicks are missing a pulse, when compared with the standard clicks that are the focus of this study. However, a cepstral analysis of this click subset shows weak peaks at 6.37 and 8.1 ms, the same IPIs present in the standard clicks. A close visual inspection of the Hilbert transforms of Fig. 5 does seem to indicate weak local maxima arriving at about 6 and 8 ms after the arrival of the intense, temporally compact main pulse at 0 ms.

There a variety of explanations as to what could be happening, but given a sample size of 1, such speculation is premature. Additional data will be needed to determine if these ambiguous clicks appear consistently during an actual depredation event.

## V. CONCLUSION

Video and audio recordings of a prey acquisition attempt from depredating sperm whales in the Gulf of Alaska have been processed to compare visual and acoustic methods for estimating animal size. The IPI between the B and C pulses (interpreted as the  $P_1$  and  $P_2$  pulses in the bent-horn hypothesis) yields a size estimate that matches visual estimates to within experimental error, a result consistent with previous empirical studies. However, the video data also permit bounds to be placed on the size of the spermaceti organ of the animal, and those results suggest that the IPI might not be a direct measure of the size of this organ. Furthermore, the size of the junk derived from the acoustic data is smaller than the estimated size of the spermaceti organ, which is inconsistent with anatomical fact.

This paper has provided a glimpse of a possible new approach for investigating the biosonar of large marine mammals in the wild, which permits close-range measurements of the acoustic structure of the terminal buzz or creak of the animal, from both a broadside and on-axis orientation that lies within the main beam of the animal. These measurement locations are unavailable to bioacoustic tags, which can only measure sounds from orientations behind the animal. Only future work will tell whether close-range observations of depredating whales can yield sample sizes sufficient to draw additional conclusions about the biosonar properties of sperm whales, and whether echoes from prey items might be identified.

## ACKNOWLEDGMENTS

The authors thank Marty, Tamantha and Jesse for allowing us to dedicate a portion of their black cod quota to the experiment. We also thank the videocamera analysis efforts of Amanda Koltz, the feedback of Ted Cranford and two anonymous reviewers, whose comments have substantially improved the paper's quality and focus. This work was supported by Grant No. 7973-06 from the National Geographic Society and North Pacific Research Board grant 626. With regard to potential regulatory issues, the authors consulted with the Permits, Conservation and Education Division, Office of Protected Resources, National Marine Fisheries Service (NMFS), which is processing the pending amendment to NMFS research Permit No. 473-1700-02, to cover future camera work that may involve more substantial modifica-

tions of longlining fishing techniques. The authors also thank the Mammalogy Collections Manager Jim Dines and the Curator Dave Janiger from the Natural History Museum of Los Angeles County as well as the Collections Manager Tony Dumitru and the Curator Ben Simons from the Nantucket Historical Association Whaling museum for their help with collecting the tooth spacing measurements.

- <sup>1</sup>W. T. Fitch, "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," *J. Acoust. Soc. Am.* **102**, 1213–1222 (1997).
- <sup>2</sup>G. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, 1960).
- <sup>3</sup>T. Nearey, *Phonetic Features for Vowels* (Indiana University Linguistics Club, Bloomington, 1979).
- <sup>4</sup>P. Lieberman, *The Biology and Evolution of Language* (Harvard University Press, Cambridge, MA, 1984).
- <sup>5</sup>G. E. Peterson and H. L. Barney, "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," *J. Acoust. Soc. Am.* **24**, 175–184 (1952).
- <sup>6</sup>H. Whitehead, *Sperm whales: Social Evolution in the Ocean* (University of Chicago Press, Chicago, IL, 2003).
- <sup>7</sup>N. Jaquet, S. Dawson, and L. Douglas, "Vocal behavior of male sperm whales: Why do they click?," *J. Acoust. Soc. Am.* **109**, 2254–2259 (2001).
- <sup>8</sup>S. L. Watwood, P. J. O. Miller, M. P. Johnson, P. T. Madsen, and P. L. Tyack, "Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*)," *J. Anim. Ecol.* **75**, 814–825 (2006).
- <sup>9</sup>P. J. O. Miller, M. P. Johnson, and P. L. Tyack, "Sperm whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey," *Proc. R. Soc. London, Ser. B* **271**, 2239–2247 (2004).
- <sup>10</sup>K. S. Norris and G. W. Harvey, "A theory for the function of the spermaceti organ of the sperm whale (*Physeter catodon* L.)," in *Animal Orientation and Navigation*, edited by S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs, and R. E. Belleville, NASA Special Publication No. **262**, Washington, DC, pp. 397–417, 1972.
- <sup>11</sup>J. C. D. Gordon, "Behavior and ecology of sperm whales of Sri Lanka," Ph.D. thesis, University of Cambridge, Cambridge, UK (1987).
- <sup>12</sup>H. Whitehead and L. Weilgart, "Patterns of visually observable behavior and vocalizations in groups of female sperm whales," *Behaviour* **118**, 275–296 (1991).
- <sup>13</sup>J. C. Goold and S. E. Jones, "Time and frequency domain characteristics of sperm whale clicks," *J. Acoust. Soc. Am.* **98**, 1279–1291 (1995).
- <sup>14</sup>A. Thode, D. K. Mellinger, S. Stienessen, A. Martinez, and K. Mullin, "Depth-dependant features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico," *J. Acoust. Soc. Am.* **116**, 245–253 (2002).
- <sup>15</sup>B. Mohl, E. Larsen, and M. Amundin, "Sperm whale size determination: Outlines of an acoustic approach," *Fisheries Series, FAO Rome*, 1981, pp. 327–332.
- <sup>16</sup>B. Mohl, "Sound transmission in the nose of the sperm whale, *Physeter Catodon*. A post mortem study," *J. Comp. Physiol.* **187**, 335–340 (2001).
- <sup>17</sup>B. Mohl, M. Wahlberg, P. T. Madsen, A. Heerfordt, and A. Lund, "The monopulsed nature of sperm whale clicks," *J. Acoust. Soc. Am.* **114**, 1143–1154 (2003).
- <sup>18</sup>W. M. X. Zimmer, P. L. Tyack, M. P. Johnson, and P. T. Madsen, "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," *J. Acoust. Soc. Am.* **118**, 3337–3345 (2005).
- <sup>19</sup>W. M. X. Zimmer, P. T. Madsen, V. Teloni, M. P. Johnson, and P. L. Tyack, "Off-axis effects on the multi-pulse structure of sperm whale usual clicks with implications for the sound production," *J. Acoust. Soc. Am.* **118**, 3337–3345 (2005).
- <sup>20</sup>V. Teloni, W. M. X. Zimmer, M. Wahlberg, and P. T. Madsen, "Consistent acoustic size estimation of sperm whales using clicks recorded from unknown aspects," *J. Cetacean Res. Manage.* **9**, 127–136 (2007).
- <sup>21</sup>N. Nishiwaki, S. Oshumi, and Y. Maeda, "Changes in form of the sperm whale accompanied with growth," *Sci. Rep. Whales Res. Inst.* **17**, 1–13 (1963).
- <sup>22</sup>M. R. Clarke, "Structure and proportions of spermaceti organ in sperm whale," *J. Mar. Biol. Assoc. U.K.* **58**, 1213–1222 (1978).
- <sup>23</sup>J. C. D. Gordon, "Evaluation of a method for determining the length of sperm whales (*Physeter catodon*), from their vocalisations," *J. Zool. (London)* **224**, 301–314 (1991).
- <sup>24</sup>M. Q. Rhineland and S. M. Dawson, "Measuring sperm whales from

- their clicks: Stability of interpulse intervals and validation that they indicate whale length," *J. Acoust. Soc. Am.* **115**, 1826–1831 (2004).
- <sup>25</sup>B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The frequency analysis of time series for echoes; cepstrum pseud-autocovariance cross-cepstrum and shape-cracking," *Symposium on Time Series Analysis*, edited by M. Rosenblatt (Wiley, New York, 1963), Vol. **15**, pp. 209–243.
- <sup>26</sup>J. C. Goold and S. E. Jones, "Sound velocity measurements in spermaceti oil under the combined influences of temperature and pressure," *Deep-Sea Res., Part I* **43**, 961–969 (1996).
- <sup>27</sup>M. Wahlberg, "The acoustic behaviour of diving sperm whales observed with a hydrophone array," *J. Exp. Mar. Biol. Ecol.* **281**, 53–62 (2002).
- <sup>28</sup>W. A. Watkins, M. A. Daher, N. A. DiMarzio, A. Samules, D. Wartok, K. M. Fristrup, P. W. Howey, and R. R. Maiefski, "Sperm whale dives tracked by radio tag telemetry," *Marine Mammal Sci.* **18**, 55–78 (2002).
- <sup>29</sup>M. R. Clarke and N. Macleod, "Cephalopod remains from sperm whales caught off Iceland," *J. Mar. Biol. Assoc. U.K.* **56**, 733–750 (1976).
- <sup>30</sup>D. W. Rice, "Sperm Whales," in *Handbook of Marine Mammals*, edited by S. H. Ridgway and R. Harrison (Academic, London, 1989), Vol. **4**, pp. 177–233.
- <sup>31</sup>J. R. Ashford, P. S. Rubilar, and A. R. Martin, "Interactions between cetaceans and longline fishery operations around South Georgia," *Marine Mammal Sci.* **12**, 452–457 (1996).
- <sup>32</sup>D. Capdeville, "Interaction of marine mammals with the longline fishery around the Kerguelen Island Division, 58.5.1 during the 1995/96 cruise," *Ccamlr Sci.* **4**, 171–174 (1997).
- <sup>33</sup>C. P. Nolan and G. M. Liddle, "Interactions between killer whales (*Orcinus orca*) and sperm whales (*Physeter macrocephalus*) with a longline fishing vessel," *Marine Mammal Sci.* **16**, 658–664 (2000).
- <sup>34</sup>E. F. Gonzalez, presented at the XXI Congreso de Ciencias del Mar, Chile (2001).
- <sup>35</sup>P. S. Hill, J. L. Laake, and E. Mitchell, "Results of a pilot program to document interactions between sperm whales and longline vessels in Alaska waters," U.S. Department of Commerce, Alaska Fisheries Science Center, Report No. NOAA, TM-NMFS-AFSC-108, 42 pp., 1999.
- <sup>36</sup>R. Huckle-Gaete, C. A. Moreno, J. Arata, and Blue Whale Ctr, "Operational interactions of sperm whales and killer whales with the Patagonian toothfish industrial fishery off Southern Chile," *Ccamlr Sci.* **11**, 127–40 (2004).
- <sup>37</sup>M. G. Purves, D. J. Agnew, E. Balguerias, C. A. Moreno, and B. Watkins, "Killer whale *Orcinus orca* and sperm whale *Physeter macrocephalus* interactions with longline vessels in the patagonian toothfish fishery at South Georgia, South Atlantic," *Ccamlr Sci.* **11**, 111–126 (2004).
- <sup>38</sup>M. F. Sigler, C. R. Lunsford, J. M. Straley, and J. B. Liddle, "Sperm whale depredation of sablefish longline gear in the northeast Pacific Ocean," *Marine Mammal Sci.* **24**, 16–27 (2008).
- <sup>39</sup>A. Thode, J. Straley, C. O. Tiemann, K. Folkert, and V. O'Connell, "Observations of potential acoustic cues that attract sperm whales to longline fishing in the Gulf of Alaska," *J. Acoust. Soc. Am.* **122**, 1265–1277 (2007).
- <sup>40</sup>J. Bouguet, Camera Calibration Toolbox for Matlab (2008).
- <sup>41</sup>T. Cranford, "The sperm whale's nose: Sexual selection on a grand scale?," *Marine Mammal Sci.* **15**, 1133–1157 (1999).
- <sup>42</sup>M. Amundin, "Sound production in odontocetes with emphasis on the harbour porpoise *Phocoena phocoena*," Ph.D thesis, University of Stockholm, Stockholm, Sweden (1991).
- <sup>43</sup>P. T. Madsen, M. Wahlberg, and B. Mohl, "Male sperm whale (*Physeter macrocephalus*) acoustics in a high latitude habitat: Implications for echolocation and communication," *Behav. Ecol. Sociobiol.* **53**, 31–41 (2002).