

Variation in humpback whale (*Megaptera novaeangliae*) song length in relation to low-frequency sound broadcasts

Kurt M. Fristrup^{a)}

Bioacoustics Research Program, Cornell Laboratory of Ornithology, Ithaca, New York 14853

Leila T. Hatch

Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853

Christopher W. Clark

Bioacoustics Research Program, Cornell Laboratory of Ornithology, Ithaca, New York 14853

(Received 27 December 2000; revised 19 February 2003; accepted 17 March 2003)

Humpback whale song lengths were measured from recordings made off the west coast of the island of Hawai'i in March 1998 in relation to acoustic broadcasts ("pings") from the U.S. Navy SURTASS Low Frequency Active sonar system. Generalized additive models were used to investigate the relationships between song length and time of year, time of day, and broadcast factors. There were significant seasonal and diurnal effects. The seasonal factor was associated with changes in the density of whales sighted near shore. The diurnal factor was associated with changes in surface social activity. Songs that ended within a few minutes of the most recent ping tended to be longer than songs sung during control periods. Many songs that were overlapped by pings, and songs that ended several minutes after the most recent ping, did not differ from songs sung in control periods. The longest songs were sung between 1 and 2 h after the last ping. Humpbacks responded to louder broadcasts with longer songs. The fraction of variation in song length that could be attributed to broadcast factors was low. Much of the variation in humpback song length remains unexplained. © 2003 Acoustical Society of America. [DOI: 10.1121/1.1573637]

PACS numbers: 43.80.Nd, 43.80.Ka, 43.30.Vh [WA]

I. INTRODUCTION

In 1997–1998, a series of experiments was undertaken to quantify the responses of selected baleen whale species to powerful, low-frequency acoustic broadcasts, including the proposed operational use of the U.S. Navy SURTASS LFA sonar system. One phase of this research focused on endangered humpback whales in Hawai'i (Clark and Tyack, 1998; Miller *et al.*, 2000). Several factors suggested that Hawaiian humpbacks were appropriate subjects for this work. The near-shore distribution of humpback whales in Hawai'i has fostered extensive research, providing an excellent record of behavior prior to the experiments. The waters around the Hawaiian Islands host high densities of breeding humpbacks and their calves, so there would be significant impact on reproductive activities if they were displaced or their behavior was seriously disrupted. Lastly, the long, elaborate songs produced by males have substantial signal energy in the range of frequencies produced by the SURTASS LFA sonar system, so humpbacks are assumed to hear and potentially be responsive to LFA signals.

Several studies have used detailed visual observation methods to investigate the responses of humpback whales to acoustic stimuli. Dramatic behavioral responses have been observed to playbacks of conspecific sounds (Tyack, 1983; Baker and Herman, 1984; Mobley *et al.*, 1988). Behavioral responses were documented in reaction to active sonar (3.1–3.6 kHz) (Maybaum, 1990, 1993). During the Acoustic Ther-

ometry of the Ocean Climate (ATOC) Marine Mammal Research Program, subtle, short-term effects on the surface behaviors of Hawaiian humpback whales were observed in response to low-frequency (75 Hz) sound broadcasts (Frankel and Clark, 1998, 2000). Although there was no decrease in humpback whale abundance in relation to broadcast activity, ATOC did reveal changes in the distribution of animals relative to the transmitter (Frankel and Clark, 2002). These results document the potential complexity of responses to a sound source. On average, the distance of pods from the transmitter increased, but the number of animals sighted near the source also increased. Miller *et al.* (2000) made focal animal behavioral observations in parallel with the data reported here. They compared the lengths of songs sung by six individuals before, during, and after exposure to low-frequency broadcasts, and found that songs were significantly longer during broadcasts.

Additional studies have documented behavioral reactions of humpback whales to vessels (reviewed in Richardson *et al.*, 1995; see also Corkeron, 1995; Frankel and Clark, 1998, 2000; Au and Green, 2000). These reactions include changes in direction and swimming speed, and changes in the pattern of surfacing. Humpback whales also modify some elements of their acoustic behavior when approached by boats: the duration of song units (notes) decreased, increasing the "tempo" of songs (Norris, 1994).

Humpback whale song is an extraordinarily long and complex acoustic display. It consists of sequences of broadband units of sound, exhibiting repetition within sequences of units and on the longer time scales of themes and songs.

^{a)}Electronic mail: kmf6@cornell.edu

The diversity of units is considerable, spanning a range of frequencies from approximately 30 to 5000 Hz. The general structure of songs sung by the majority of males at any one time and place is similar (Payne and McVay, 1971; Cerchio *et al.*, 2001). However, the detailed structure of successive songs sung by an individual varies (Payne *et al.*, 1983; Payne and Payne, 1985; Helweg *et al.*, 1992). Variation in the structure of an individual's song may relate to interactions within a humpback chorus.

Studies have shown that singers are male, and that singing is mainly associated with low-latitude, coastal areas where calves are born and mating is presumed to occur. The function of humpback whale song is disputed, but hypotheses have focused on its probable reproductive context (Payne and McVay, 1971; Clapham, 1996), and have sought analogs in the mating systems of birds, frogs, and insects. As has been well documented in many terrestrial systems, humpback song may mediate interactions among males (Tyack, 1981; Darling, 1983; Frankel *et al.*, 1995), and advertise species, gender, location, and condition to females (Payne and McVay, 1971; Winn and Winn, 1978; Tyack, 1981).

Any analysis of response to potentially aversive stimuli must incorporate provisions for differences among individuals, and for the variation in each individual's behavior. With respect to humpback song in particular, differences among individuals are expected if song plays a role in female mate choice or mutual assessment of competitive ability among males. Aspects of acoustic displays that reveal differences among individuals have been shown to evoke predictable female responses in insect, frog, and bird species (*e.g.*, Catchpole, 1980; Klump and Gerhardt, 1987; Eiriksson, 1994; Brown *et al.*, 1996; Welch *et al.*, 1998; Gentner and Hulse, 2000).

Humpback song length is a simple summary of a complex behavior that is likely to provide an informative measure of response to LF broadcasts for several reasons. As an indicator of the regularity and rhythm of display behavior, song length provides a relatively easily extracted measure of response to potential disturbance. Also, it has been argued that the consistent production of longer humpback songs is a reliable indication of superior condition, due to the constraints that longer songs place on respiration (Chu and Harcourt, 1986; Chu, 1988). It also seems plausible that longer songs imply greater energetic investment in a "unit" of display. Although a complete humpback song may not be analogous to a single frog call, Taigen and Wells (1985) demonstrated that female frogs were more attracted to longer calls, which were more energetically expensive. However, Helweg *et al.* (1992) theorized that humpback song comprises a negligible fraction of their energy budget, which suggests that physiological constraints on song length and loudness are not likely to be related to energetics.

Data are presented on the lengths of 378 humpback songs recorded before, during, and after low-frequency acoustic broadcasts. Statistical models of song length in relation to a variety of natural and experimental factors were developed. Song length was analyzed in relation to date and time of the song, the identity of the singer, and several factors related to the acoustic broadcast. Song length variability

during control periods was documented to provide a context for assessing the biological significance of changes related to low-frequency broadcasts.

II. METHODS

These data were collected from 26 February to 29 March 1998 off the western coast of the island of Hawai'i, between Mahaiula and Kawaihae. The study area was chosen to utilize extensive baseline data on the abundance, densities, movements, and associations of humpback whales (Gabriele, 1992; Craig and Herman, 1997; Frankel *et al.*, 1995; Mobley *et al.*, 1994, 1995, 1999). Several of these studies showed that the majority of whales were found within the 100-fathom contour, though the fraction of singers farther offshore has been as high as 30% (Herman and Antinaja, 1977; Frankel *et al.*, 1995). Photo identification studies showed that most whales remained in the area for 4 to 7 days (Gabriele, 1992). This short residency time suited the objectives of this study by limiting the number of times any one animal would be exposed to the experimental sounds. The study was conducted after the peak in seasonal abundance, because previously the research team was conducting related research during the peak of gray whale migration off Central California.

This research used U.S. Navy SURTASS LFA sound projectors to broadcast low-frequency ("LF") sounds in the 150–320-Hz frequency band. An LF signal (or "ping," a bit of sonar jargon borrowed for brevity) consisted of nine sound units lasting a total of 36 s and spanning a total of 42 s. Two types of pings were alternated, a "high" ping (260–320 Hz) and a "low" ping (150–230 Hz). Each day an initial series of reduced amplitude pings were transmitted to test the equipment and to allow animals to become aware of the sound source before full experimental amplitude was reached. The first ping of each day was transmitted at a source level (SL) of 155 dB *re* 1 μ Pa at 1 m (in the remainder of the paper, dB used). SL increased progressively until the predicted received level reached the desired value. During an experiment, pings were broadcast in a series of ten, on intervals of 6 min (see Fig. 1), for a total of 54 min in a ping series. SL was monitored via a calibrated hydrophone array system, and did not exceed 205 dB. Source levels were adjusted to realize maximum received levels of between 120 and 155 dB *re* 1 μ Pa at the nearest whale, while ensuring that exposure to animals within 3 miles of shore and to human divers did not exceed 140 dB *re* 1 μ Pa. These broadcasts achieved the same range of received levels that over 95% of the animals are predicted to experience during the operational training exercises for this equipment (Navy, 2001, appendix D). These received levels were accomplished, in spite of lower source levels, by closely approaching the experimental subjects. This procedure minimized exposure levels to animals not under observation. Ping series were separated by at least 2 to 3 hours. Up to three ping series were produced each day.

Data were collected regarding the short-term responses of individual whales, and longer-term changes in the distribution and sighting rates of animals in the area. Two methods

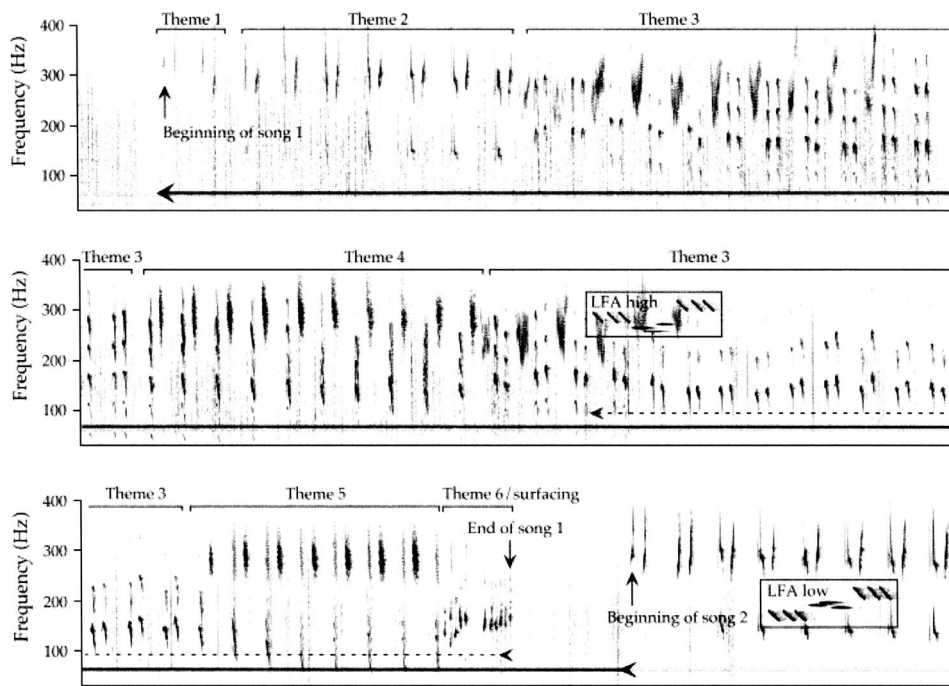


FIG. 1. Spectrogram of humpback song and two pings, with themes and song length measurements marked. Each panel represents five minutes of sound. The solid line marks the measurement of song length. The dotted line marks the measurement of minutes from the start of the last ping to the end of the song.

were used to follow the short-term behavior of individual singers. Visual and acoustic methods were used to locate and follow humpback singers from a small inflatable boat (Miller *et al.*, 2000). If the visual observers selected a focal animal, its behaviors were recorded for at least two dive cycles before the SURTASS LFA vessel moved toward the singer and broadcast sounds. In parallel, acoustic observers on the SURTASS LFA vessel collected the data reported here. Beam-forming software was used to locate humpback singers who were candidates for broadcast experiments. If the acoustic observers selected a focal animal, baseline acoustic behavior was recorded for at least three complete songs before a ping series began.

Long-term changes in distribution and abundance were monitored using visual surveys from a shore station and from the SURTASS vessel. Shore station observations followed standardized protocols (Frankel *et al.*, 1995; Frankel and Clark, 1998, 2000). SURTASS vessel observations followed a protocol adapted from established visual survey techniques (Barlow, 1995). Five observers rotated through four stations (three observing, one data recording) on half-hour intervals throughout the day. A senior National Marine Fisheries Service (NMFS) observer continuously oversaw this effort. A portion of these data is used here to indicate possible social contexts for the acoustical patterns exposed by the analysis.

The acoustic behaviors of whales in the vicinity (<4 km) of the playback vessel were monitored and recorded using a long, low-frequency, horizontal hydrophone array towed behind the playback vessel. Fifteen elements in the array were used to collect time series data. The acoustic data collection system operated from 0400 h until approximately 1800 h local time each day, although data collection effort often continued throughout the night. The data were recorded digitally on a Windows 98 computer at a sampling rate of 1002 Hz per hydrophone.

These array data were intensively analyzed to identify

singers and follow their acoustic behavior, focusing on the song duration. The songs of individual singers were followed for as long as possible. The ability to keep track of individual singers relied on the stereotyped structure of the humpback whale song (Fig. 1), and spatial cues provided by the array recordings (relative intensities, time delays). For the purpose of these analyses, two themes of humpback song were relevant. A series of trills, also called the “ratchet,” was designated theme 6. It was typically followed by a short period with no acoustic activity between 0 and 500 Hz. This hiatus in low-frequency activity was followed by a series of frequency-modulated (FM) upsweeps, which was designated theme 1. Theme 6 has been observed across many seasons and populations, and has typically preceded a respiratory surfacing. Although singing whales do not surface exclusively at this point in the song, the end of this theme has previously been referred to as the end of a song (Winn and Winn, 1978; Cerchio *et al.*, 2001).

A data visualization program written in MATLAB (The Mathworks, Inc., 1999) enabled discrimination among different humpback singers and facilitated measurement of humpback song duration. An operator used the program to transcribe acoustic data into information about the beginning and ending song units, including the identity of the singer and the temporal and frequency bounds of the units. Four spectrogram windows, representing 2.5 min of sound from the selected channel, displayed a total of 10 min of contiguous sound per screen. The temporal extent of this display permitted viewing most of a song while providing sufficient resolution to see structure in song units. A point in the song could be selected with a cursor, and two other windows displayed an expanded view of 20 s of data from two channels of the array.

Although each broadcast experiment sought to isolate single singers within a kilometer of the vessel, several singers were usually detectable within the array’s acoustic range.

Difficulties in following the thematic structure of low-amplitude songs, or songs in dense choruses, limited the conditions under which a sequence of humpback sounds could be unambiguously attributed to one animal. The songs of an isolated, loud, slow moving singer could be followed for many songs, but most singers were followed for a few songs.

A continuous sequence of sounds could be attributed to one singer using a combination of cues: continuity of song units and themes, the pattern of received levels across the array, and the pattern of arrival time delays across the array. Song length was measured as the interval between successive starts of theme 1 or successive endings of theme 6, whichever was clearer (Fig. 1). Note that this measurement is unambiguous, even in this example, which Frumhoff (1983) would term an "aberrant" song. The continuity of a singer's acoustic behavior became ambiguous when (1) there was a break of over 2.5 min (one browser panel) in the middle of a song (due to changes in detection and/or singing behavior), and (2) when multiple singers were at similar distances and similar bearings and were singing the same theme. Both cases produced uncertainty regarding singer identity, and subsequent songs were attributed to a new singer. Note that some individuals were sampled more than once, so the number of singers in these analyses overstates the number of individuals.

These measurements of song duration were analyzed in relation to several factors that could systematically affect singing behavior. These factors were day of year (1 = 1 January 1998), time of day (00:00:00–23:59:59), singer identification number, minutes since last ping, minutes since the beginning of a ping series, minutes of pings during a song, and LF source level. The distinction between the number of minutes since the last ping and the number of minutes since the beginning of a ping series was used to investigate a potential cumulative effect. Figure 1 provides an example of the measurements taken on song duration and the timing relationship measured between songs and pings.

Two statistical procedures were used. The simplest assessment of playback effects was to examine differences in song length in relation to discrete temporal categories. Miller *et al.* (2000) defined temporal categories based on *a priori* assumptions regarding behavioral response: before ping series, during ping series, after ping series. The analysis presented here identified temporal categories that produced maximal contrasts in song length. These categories were identified using a tree-based regression (Chambers and Hastie, 1991) of song length on the minutes elapsed since the last ping. The six-category tree provides a detailed picture of potential responses while retaining reasonable sample sizes within categories. The significance of differences among these categories was measured by pairwise *t*-tests; no correction was made for multiple inferences. Differences in song length were also examined in relation to the amount of overlap by pings.

These simple analyses ignored the potential confounding effects of other factors. Three accessible factors seemed important to incorporate into a model of response: date (seasonal effects), time of day (diurnal effects), and ping source level. Source level was incorporated for two reasons. It

seemed plausible that whales have evolved the capacity to gauge the source levels of nearby sounds, especially if song source level provides important clues to the singers' status. Second, source level served as a proxy for received level, with the caution that extensive variation in transmission loss could cause source and received levels to be weakly correlated.

It would be preferable to analyze and compare models using both source and received levels, but significant obstacles remain. Receiver depth is the critical factor affecting received levels, especially when the source to receiver distance exceeds 1 km (as it did in virtually all of these experiments). Matched-field processing has been used to infer the depths of singing blue whales (Thode *et al.*, 2000), and ongoing research may yield fruitful methods for these data. However, the subsequent analysis of response would be much more complex than the material presented here. Unlike source level, received level varies significantly during and among pings, as the animal moves and changes depth. Alternative methods of summarizing each subject's received level history would need to be developed and tested to identify summary values that provide the best predictors in a response model.

Multivariate general linear models do not allow for the nonlinear pattern revealed by the categorical analysis, so generalized additive models (GAMs) were used (Hastie and Tibshirani, 1998). Generalized additive models are analogs of linear statistical models, in which the effects of factors are represented by nonparametric smooth curves. These smooth curves provide estimates of the local average of the data, extending the notion of categorical averages to a continuous representation. The models used here employed smoothing splines to estimate the nonlinear effects of all the factors that, when added together, provided the best fit to the observed data. These splines were fitted using an iterative backfitting routine, whose convergence properties are specified by analyses of the Gauss–Seidel algorithm (Hastie and Tibshirani, 1998).

The divergence of these splines from linear models was constrained by specifying the equivalent degrees of freedom used in their computation. Hastie and Tibshirani (1998) discuss the theoretical bases for calculating degrees of freedom for smoothing functions. The extent of the data used to estimate the local averages is inversely proportional to the degrees of freedom used in fitting the spline. As more degrees of freedom are used, the bias of local estimates decreases at the cost of increasing the variance of the estimates (and the confidence intervals for prediction). This bias-variance tradeoff is unavoidable. Three degrees of freedom were used for all smoothing splines in these models, with the exceptions that the factor representing the minutes elapsed since the last ping used 6 degrees of freedom, and the factor representing the minutes of ping overlap with each song used 2 degrees of freedom. The goal in all cases was to achieve a relatively uniform distribution of residual errors while consuming the fewest degrees of freedom.

The significance of a fit between the smooth trends and the data was assessed by relating the amount of variation explained to the degrees of freedom in the model. *F*-ratio

tests were used to test for the significance of contributions from the different factors. All analyses were performed in S-Plus (Statistical Sciences, Inc., 2000).

In addition to the minutes elapsed since the last ping (“single ping” models), two other factors were used to investigate potential temporal patterns in the responses to LF broadcasts. The minutes elapsed since the beginning of a ping series was included to see if the effects of exposure to a series of pings were cumulative (“ping series” models). The total number of minutes of pings that overlapped each song was included to see if males reacted to the amount of potential interference during each song (“overlap” models).

The categorical analyses and generalized additive models pooled data across song series and across days and times of day to utilize as many measurements as possible to estimate response. Contrasts could have been formed between the preexposure, experimental, and postexposure behaviors within each song series. However, it was difficult to obtain a full matched sample of data in a song series. This “matched sample” approach would have been limited to a maximum of 15 song series in this data set; these limited data would not support models testing the effects of diurnal and seasonal effects.

By pooling the data, song length measurements for 113 singers were available for analysis in the models incorporating date and time of day. As noted earlier, songs from a single individual may have been labeled with a succession of singer numbers, and there is no way to determine exactly how many individuals were studied. The maximum number of simultaneous singers gives a lower bound for the number of individuals present on each day. The sum of these maxima for all days was 60. Some of these animals represent repeat sightings across days. During this project, Biassoni (2000) found that 3 of 23 animals selected for focal behavioral studies had been selected as the focal animal previously. The upper 95% confidence interval for this proportion is about 0.25. Thus, a conservative lower bound on the number of individuals sampled for this study is 45.

The magnitude of individual variation in song length was investigated by examining the residuals of the multivariate GAM that fitted time of day, day of year, minutes since last ping, and source level to the song length measurements. After removing the variation in song length that can plausibly be attributed to generic factors, this analysis should provide a reliable indication of the extent of variation among individuals. Song series consisting of five or more songs were used, resulting in a sample of 25 song series comprising 189 songs. A tree-based regression was utilized to identify homogeneous groupings of song series, and a Kruskal–Wallis test was used to assess the significance of differences among these groupings.

Although the song series measurements were not long enough to enable simultaneous estimation of singer idiosyncrasies with time of day and day of year effect in a GAM, a second group of GAM analyses were computed replacing the generic diurnal and seasonal factors with a factor that allowed each song series to have a different average value. For brevity, these models will be called “singer ID” models. Comparison of these singer ID models with the generic mod-

els was used to assess how consistent the estimated effects of the LF broadcasts were. Song series containing at least three song measurements were used for the singer ID models, which restricted the data set to the behavior of 52 song series comprising 290 songs.

The contributions of each smoothing factor are plotted as smooth curves against a background that depicts the range of variation in song length. These smooth curves illustrate the song lengths predicted by each factor, assuming all other factors make an average contribution. A series of curves are presented for each factor, illustrating the results obtained from different generalized additive models. With the exception of Fig. 10, the null hypothesis of no effect corresponds to a horizontal line in these figures.

For Fig. 10, the null hypothesis must reflect the fact that if whales were oblivious to the pings, then minutes of overlap should be proportional to song length. In particular, a song of any length could be overlapped by a single ping, but songs overlapped by two pings would have to be more than 6 min long, and songs overlapped by three pings would have to be more than 12 min long. The appropriate null hypothesis—no reaction to the pings—thus corresponds to a linear relationship between song length and overlap with a slope equal to the inverse of the ping duty cycle ($6 \times 60 \text{ s}/42 \text{ s}$).

The plots of smoothing factors (Figs. 8, 9, and 11) contain three indications of the extent of variation in humpback song length. A background scatter plot shows the humpback song length measurements against the values of the factor. A background histogram shows the distribution of lengths for songs sung before the first ping of each day. More than 12 h elapsed between the last ping on the previous day and any of these songs, and analyses presented below indicate that responses to LFA signals disappear 2 h after the last ping. However, these songs do not represent a balanced sample with respect to diurnal factors. The early morning is disproportionately represented, so average differences between these control song lengths and other samples may include diurnal effects. The third measure of variation is represented by a dark bar near the vertical axes whose length depicts 2.5 min, which is the average difference between the lengths of successive songs sung by an individual. These graphic displays of variation provide a natural scale for interpreting the magnitude of fitted effects.

Three exceptional songs exceeded 30 min in length; the next longest song was 22.9 min long. Two of these songs were sung consecutively by one individual, starting at 1500 h, 13 min after the end of a ping series, 24 days after the LF broadcast experiments began (4 days before they concluded). The third song was sung the following day, at 1639 h, 45 min after the end of a ping series. These songs were excluded from all statistical models to avoid disproportionate influence on the results.

III. RESULTS

Array recordings were collected in conjunction with LF broadcast experiments from 2 to 29 March 1998. Song length measurements were obtained by browsing 121 h of data distributed across 23 days. No experiments were per-

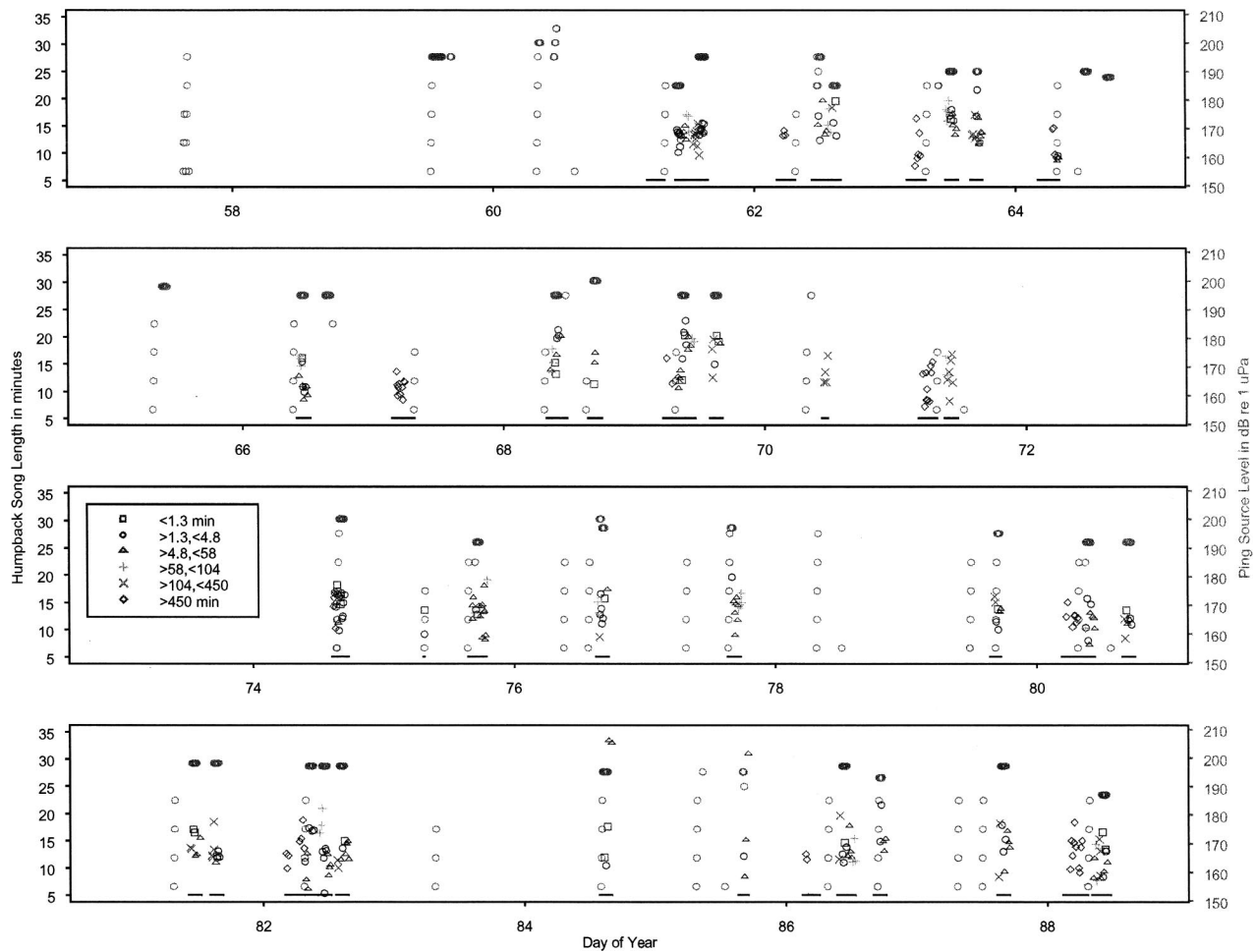


FIG. 2. History of humpback song length in relation to ping source level. Song length measurements are represented by one of six symbols, which show the categories from the analysis presented in Fig. 3. The gray circles show the time and source level of each ping. The solid lines near the bottom of each panel show the extent of acoustic sampling effort.

formed on 7 days due to high winds and other impediments to operations. Humpback song length was highly variable. Mean song length was 13.8 min (s.d.=3.1, minimum=5.4, 1st quartile=11.7, median=13.5, 3rd quartile=15.5, maximum=33.3 min, $N=378$). The difference between the shortest and longest songs measured on all days exceeded 10 min. The average difference in the lengths of successive songs sung by a singer was about 2.5 min. The average standard deviation for a series of songs sung by an individual was 2.76 min ($N=341$). These measures of intrinsic variation provide an important basis for assessing the scale of response. These measures consistently indicate that the variation among songs sung by an individual constitutes most of the pooled variation in song length.

Figure 2 graphically displays the time course of the project at a glance. The durations of 378 songs are displayed against the source levels of the pings and the temporal extent of the acoustic data analysis. Each song length measurement is coded by a symbol indicating its assignment into one of six categories, based on the tree-based regression of song length on minutes since the last ping. “Minutes since the last ping” were measured as illustrated in Fig. 1. These categories were: less than 1.3, 1.3 to 4.8, 4.8 to 58.5, 58.5 to 104, 104 to 450, and more than 450 min since the last ping. The

relatively even distribution of these symbols across all ping series supports the conclusion that the differences found among these categories are not potentially influenced by correlation with diurnal or seasonal factors, or the idiosyncrasies of a few trials.

Humpback whale songs that were overlapped by pings were longer than songs that were not overlapped. A Student's t test indicates that this difference is on the border of the conventional test for significance ($t=1.961$, $df=373$, p -value=0.0506), while a Wilcoxon rank sum test does not yield as strong a result ($Z=1.6368$, p -value=0.1017). Although it is conceivable that songs of less than 6 min in length could fall between pings, and thus bias this result by their inclusion in the zero ping overlap category, no such songs were observed. However, these data cannot distinguish between the effects of overlap and immediacy of the last ping, because these factors are highly correlated. The sample with zero overlap contains only two songs that ended within 12 min of the last ping, and only two songs that were overlapped ended more than 12 min after the last ping.

To illustrate the differences in song length predicted by the minutes since the last ping, Fig. 3 illustrates boxplots of the song length data for the divergent subsets identified by the tree-based regression. These boxplots illustrate the loca-

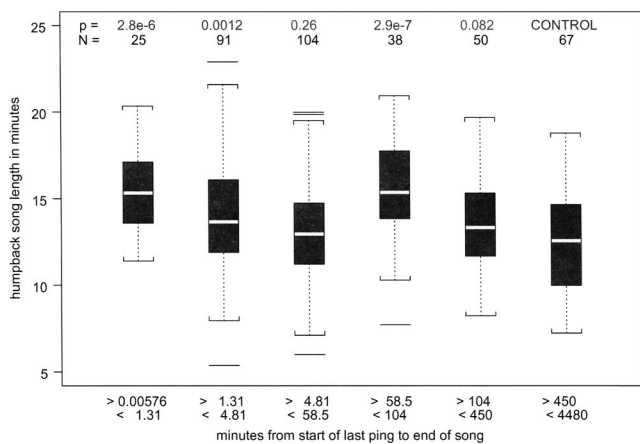


FIG. 3. Boxplots of humpback song lengths in relation to LF broadcasts. These boxplots depict distributions of song lengths in the six most distinct subsets of data, grouped by the minutes elapsed since the last ping. The solid bars indicate the range of song length values between the 25% and 75% percentile order statistics (the quartiles). The white line within the solid bars indicates the median value. The square brackets indicate the range of song length values or the range limit. The range limit was defined as one and a half times the interquartile range beyond the quartile. The horizontal bars mark measurements that fell beyond the range limit (outliers). The divisions in the data were selected by a tree-based regression of song length on the minutes from the start of the last ping, to maximize contrasts. The probability (t -test) that the subsample is drawn from the same distribution as the rightmost subsample is shown at the top, along with the sample size.

tion, scale, and asymmetry of each set of data; boxplots also mark clear outliers in the data. Songs in the last category, to the far right in Fig. 3, can be interpreted as behavior during control periods. These were recorded between 11 and 73 h after the last ping. As noted previously, these data do not provide an ideal control because they are not evenly distributed throughout the day.

Two subsets of data in Fig. 3 were not significantly different from the “control” subset. Songs that ended 104–450 min after a ping were slightly longer than those in the control subset, but overlapped sufficiently to fall short of the 5% criterion for statistical significance. Although it is possible that some of this difference is due to a lingering effect of LF broadcasts, the songs in this subset were recorded significantly later in the day than the songs in the control subset (Wilcoxon rank-sum test $Z = -7.9753$, p -value=0). A significant diurnal trend for song length is documented in the multivariate analyses below.

The third subset, songs that ended 4.8–58.5 min after the last ping, provides a counterintuitive result. This result shows that songs overlapped by a ping more than 5 min from their endings had the same distribution of lengths as all of the songs that ended within an hour of the last broadcast. About 40% of the songs in this sample ($N=42$ of 104) were overlapped. The mean length of these overlapped songs was not different from the controls ($t=0.7102$, $df=107$, p -value=0.4791) or the nonoverlapped songs in this subset ($t=-0.3116$, $df=102$, p -value=0.756).

Immediate and delayed responses were revealed by the remaining categories. Humpback songs were significantly longer when a ping occurred close the end of the song (leftmost category in Fig. 3). The magnitude of the song length increase diminished as the time since last ping increased.

This pattern may indicate that humpback whales only changed the song they were singing when a ping occurred late in a long song. Alternatively, this pattern could indicate that humpbacks increased their song length in proportion to the number of pings that overlapped their song. Pings were 6 min apart during the ping series. A song that had a ping near its end might have had one or even two ping overlaps earlier in the song.

The delayed response is represented by the fourth boxplot in Fig. 3. Songs that ended between 58.5 and 104 min after a series of pings (median length: 15.4 min) had the largest increase in mean song length relative to the control subset (median length 12.6 min). Given the documented lack of response for songs that ended between 4.8 and 58.5 min after the last LF broadcast, this delayed response is surprising. It is unlikely that this result is an artifact of an idiosyncratic distribution of these samples in relation to other factors. This sample of 38 songs is drawn from a wide range of days and times of day (Fig. 2). Additional support for this observation arises from detailed examination of the 13-song series containing at least three song length measurements ($N=3,4,4,4,5,7,7,8,8,9,10,11,13$), with at least one of these measurements between 58.5 and 104 min after the last ping. Six of these series had their maximum song length between 58.5 and 104 min after the last ping, and five others had long songs in this interval that were nearly equal to the maximum for each series.

A simple test for the effect of overlap is to compare the average lengths of songs that were overlapped versus songs that were not overlapped by pings. The mean length of the overlapped songs was slightly greater (14 vs 13.3 min), and a t -test indicates that this difference is statistically significant ($t = -1.961$, $df = 373$, p -value=0.0506). However, chance alone would cause longer songs to be more often (and extensively) overlapped by pings, if whales were oblivious to these LF signals. The “oblivious hypothesis” predicts that an increase in song length of 6 min would result in one additional overlapping ping (42 s) for songs sung entirely within the bounds of a ping series. For songs that do not fall entirely within the span of a ping series, the expected amount of overlap would still rise with increasing song length.

Figure 4 exhibits the relationship between song length and the minutes of ping overlap per song, showing that song length increases with increasing amounts of ping overlap. The three solid gray lines have a slope equal to the inverse of the duty cycle (“oblivious hypothesis”) and are drawn to pass through the mean song lengths for songs overlapped by one, two, and three pings. The oblivious hypothesis predicts much steeper increases in song length, as a function of overlap, than was observed. These deviations from the prediction are statistically significant (t -test results presented in Fig. 4). The three dotted gray lines in Fig. 4 pass through the same mean values, and have a slope equal to one. They correspond to a simple form of compensation in which whales increase the length of their songs by the amount of ping overlap. This form of compensation is also inconsistent with the observations.

Both of these univariate analyses neglected differences among singers, and the data in one or more categories may

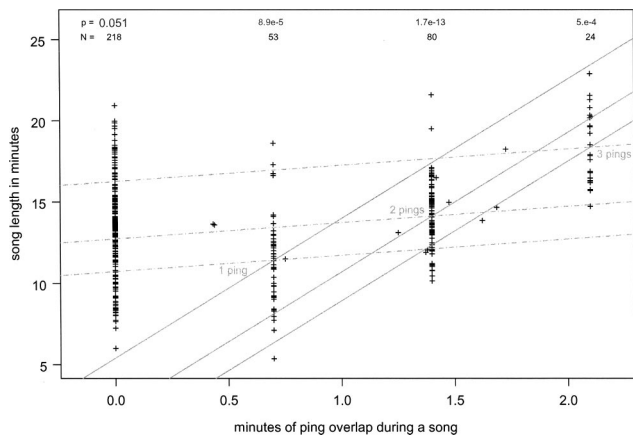


FIG. 4. Song length in relation to ping overlap. This scatterplot depicts the lengths of humpback songs in relation to the minutes of ping overlap. The sample sizes for each grouping of points are given above the clusters representing exactly zero, one, two, and three pings overlapping the songs (no fractions of a ping). The p -value in the upper left corner (0.051) is derived from a t -test comparing the songs with zero overlap with the songs that were overlapped by pings. The second p -value ($8.9e-5$) is derived from a t -test comparing the zero- and one-ping categories. The three solid lines are drawn through the median values for three categories of overlap (one, two, three pings), with a slope equal to the inverse of the duty cycle (the oblivious hypothesis). The third and fourth p -values compare the respective categories of song length measurements against the previous category, after adjusting for the slope of the solid lines. The dotted lines are drawn through the same median values, but have a slope of one (assumes songs are lengthened by the amount of time that they are overlapped).

represent skewed distributions for other factors such as date and time of day. To account for the effects of these factors simultaneously, generalized additive models were fitted. A total of 369 song length measurements from 113 song sessions were used in these analyses. Song length was fitted to day of year, hour of day, source level of the last ping, and a factor involving the timing of the broadcasts. Three broadcast timing factors were used: minutes since the start of the last ping (log 10 transformed, in the “single-ping model”), minutes since the last ping series began (log 10 transformed, in the “ping-series model”), and minutes of pings overlapping each song (in the “overlap model”).

TABLE I. Statistical results of the generalized additive models. The results on the left pertain to the generic models using time of day and day of years; the results on the right pertain to the singer ID models. Each row presents the statistical significance attributed to each factor by the different models. There are no corrections for multiple inferences in these calculations.

	Date and hour of day models			Singer models		
	Min. since last ping	Min. since last series	Min. of overlap	Min. since last ping	Min. since last series	Min. of overlap
s(day of years)	$F = 3.638\ 278$	$F = 5.067\ 14$	$F = 2.3724$			
nonparametric df=3	$p = 0.013\ 088\ 27$	$p = 0.001\ 909\ 5$	$p = 0.070\ 135\ 57$			
s(hour of day)	$F = 2.381\ 553$	$F = 3.895\ 104$	$F = 3.565\ 49$			
nonparametric df=4	$p = 0.051\ 337\ 87$	$p = 0.004\ 168\ 2$	$p = 0.007\ 238\ 68$			
s(source level)	$F = 7.212\ 696$	$F = 1.745\ 241$	$F = 6.309\ 77$	$F = 0.856\ 559$	$F = 2.818\ 95$	$F = 0.843\ 97$
nonparametric df=1	$p = 0.007\ 608\ 76$	$p = 0.187\ 369\ 8$	$p = 0.012\ 483\ 63$	$p = 0.355\ 808\ 8$	$p = 0.094\ 571\ 9$	$p = 0.359\ 339\ 2$
s(log 10(min since last ping))	$F = 4.187\ 802$			$F = 3.343\ 22$		
nonparametric df=7	$p = 0.000\ 190\ 69$			$p = 0.002\ 096\ 1$		
s(log 10(min since last series))		$F = 1.484\ 326$			$F = 1.347\ 82$	
nonparametric df=5		$p = 0.171\ 76$			$p = 0.229\ 102\ 3$	
s(pings per song)			$F = 4.03E+01$			$F = 18.867\ 39$
nonparametric df=2			$p = 2.22E-16$			$p = 2.68E-08$
null deviance/d.f.	3495/362	3458/358	2308/348	2636/280	2597/276	2636/280
residual deviance/d.f.	2687/343	2756/339	2308/348	1303/219	1351/217	1176/224

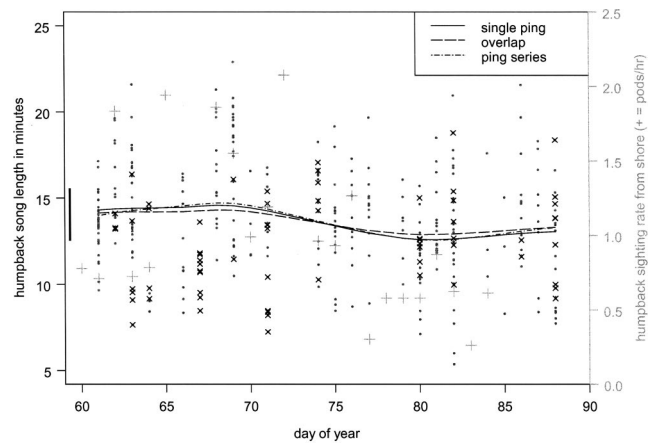


FIG. 5. Seasonal trend in song length and pod sighting rates from 1–28 March 1998. The plus symbols are normalized pod sighting rates from a shore station several miles from the site of the experiments. The small gray dot symbols indicate the song length measurements. The length of the solid bar near the left-hand axis represents the average change in the lengths of successive songs sung by a singer. The solid and dashed curves represent the smoothing splines for day of year in three generalized additive models (GAMs). These GAMs differ by the substitution of minutes from the start of last ping series or ping overlap for minutes from the start of the last ping (response to ping series or overlap vs. single pings).

The statistical results of these multivariate models are summarized in Table I. The results on the left pertain to the generic models using time of day and day of year; the results on the right pertain to the singer ID models. Each row in Table I presents the statistical significance attributed to each factor by the different models. Although there are no corrections for multiple inferences in these calculations (e.g., Bonferroni), consistently significant results provide broader support for attributing behavioral meaning to a factor.

All three models produced estimates of seasonal and diurnal factors that were remarkably consistent (Figs. 5 and 6). These figures present the fitted effects as lines, with a scatter plot of the raw data in the background. Day of year (Fig. 5) shows a modest effect, which roughly corresponded to the density of animals seen from a nearby shore station. Time of

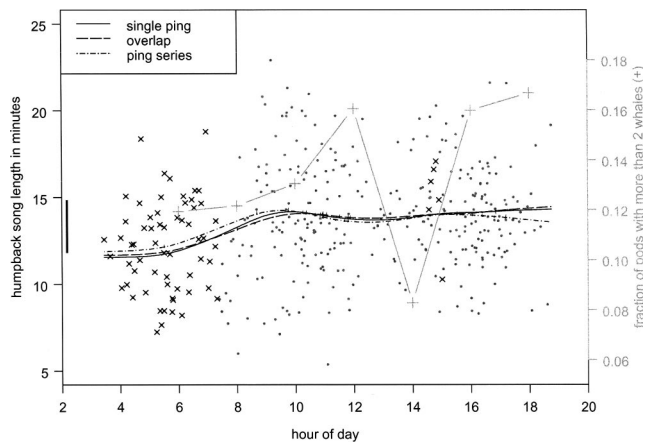


FIG. 6. Diurnal trend in song length and index of social activity. The small gray dots indicate the song length measurements. The “x” symbols indicate the lengths of songs sung before the first ping of the day. The length of the solid bar near the left-hand axis represents the average change in the lengths of successive songs sung by a singer. The plus symbols represent an index of social activity: the fraction of pods with more than two animals as seen from the source vessel. The solid and dashed curves represent the smoothing splines for the time of day factor in the same GAMs as Fig. 4.

day (Fig. 6) shows a stronger effect: songs were much shorter in the early morning. The diurnal trend in song length roughly parallels an index of social activity. These seasonal and diurnal factors made significant contributions as measured by F -ratio comparisons of the models without and with each factor (Table I, first three models). However, the ranges of these factors’ effects were approximately equal to the average difference in the lengths of successive songs sung by an individual (dark vertical bar on the left in Figs. 5 and 6).

It may appear reasonable to assert that the diurnal pattern reflects a cumulative response to LF broadcasts, because the earliest broadcasts started at 0730. The data suggest, however, that this diurnal pattern was unchanged on days when no broadcasts took place in the morning (Fig. 6). For songs sung more than 450 min after the last LF transmission, songs sung before 0730 were significantly shorter than those sung after 1400 ($t = -2.3683$, $df = 65$, p -value = 0.0209). A second caution concerns the apparent slump in song length and singing activity in the early afternoon. The apparent decrease in the number of songs (and possibly song length) in the early afternoon is largely due to sparse acoustic data analysis effort for those hours. However, the index of social activity is supported by consistent sampling effort.

Before presenting the fitted effects of the broadcast factors, a modest digression is warranted to investigate the differences in singer song length tendencies. The models discussed above neglected idiosyncratic differences among singers in order to use as many song length measurements as possible. Many song series were relatively short: 1 song length measurement was obtained from 38 song series, 2 measurements from 24 song series, 3 from 10, 4 from 16, 5 from 5, 6 from 5, 7 from 2, 8 from 5, 9 from 4, 10 from 2, 11 from 1, and 13 from one song series. A univariate analysis of song series measurements may tend to overestimate individual variation. The contributions of diurnal and seasonal factors should be factored out in order to ensure that they do not inflate the apparent differences among singers. Thus, the

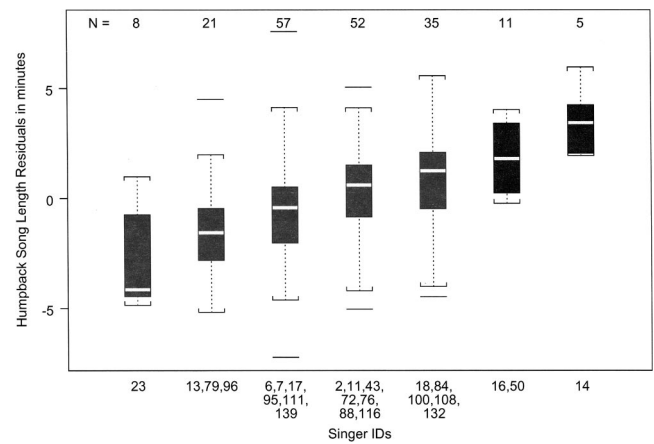


FIG. 7. Boxplots of song length residuals for groups of singers. The residual errors from model 1 in Table I were fitted using a tree-based regression on singer ID (categorical variables). The number of songs in each group are at the top of the plot. The horizontal axis labels list the singer IDs in each group. Singer IDs were assigned chronologically in this study.

residuals of the single ping model (Table I, first model) were analyzed for all song series containing more than four songs ($N = 25$ series). There are significant differences in song length among singers. A tree-based regression of song length residuals on singer ID produced seven distinct clusters (Fig. 7). Differences among all clusters are highly significant (Kruskal-Wallis chi-square = 42.775, $df = 6$, p -value = 1.29×10^{-7}). The two most extreme clusters represent single singers. Singer IDs were numbered consecutively during the study. The random distribution of the singer IDs on the horizontal axis indicates that singer idiosyncrasies did not exhibit a generic seasonal pattern.

In order to control for differences among singers in the multivariate analyses, additional GAMs were fitted by substituting a song series factor for the time of day and day of year factors. These models were compared with the previous generic time/date models to assess the stability of the fitted shapes of the broadcast factors with respect to changes in other fitted factors. Song series were too brief to investigate diurnal and seasonal trends for each singer, so in these models the singer ID term captured both temporal and individual sources of variation. These singer ID models used measurements from song series with three or more songs (52 series, 290 measurements). The results in Table I (models 4–6) indicate that the single-ping and overlap factors remained significant in the singer ID models, but the ping-series factor did not.

Figure 8 reinforces the conclusion that the ping series factor does not consistently predict song length in these models. In Table I the F -ratio tests for the second and fifth models indicate that the ping series factor does not explain a significant amount of variation in song length in either the generic or singer ID models. Figure 8 shows that the shape of the fitted curve changes dramatically when a singer ID factor is substituted for the date and time factors. In tandem, these indicate that ping series is not as good a predictor as single-ping and overlap factors. Again, the data on the extreme right provide an estimate of baseline behavior, though

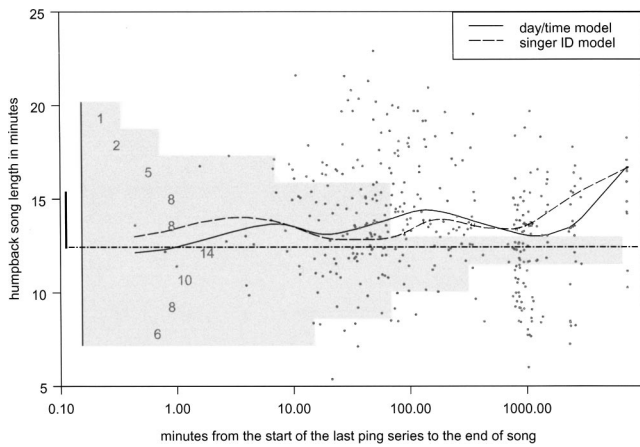


FIG. 8. The effect of time since the last ping series. The solid and dashed curves represent the smoothing spline for the ping-series factor. The solid curve is from the GAM using day of year and time of day; the dashed curve is from the GAM using individual ID. The light gray dots provide a scatterplot of the data. The underlying histogram indicates the distribution of song lengths during morning periods before the first transmission of the day. The length of the solid bar near the left-hand axis represents the average change in the lengths of successive songs sung by a singer.

songs sung in the early morning are disproportionately represented.

The multivariate single-ping model (Fig. 9) confirms the univariate results presented in Fig. 3. Baseline behavior is represented on the extreme right. Both the immediate and delayed increases in song length are evident in the multivariate models. Furthermore, the generic and singer ID models match quite closely for songs sung up to 100 min after the last ping. This indicates that the shape of this fitted factor reflects a salient relationship between the time since the last ping and song length, and not artifacts of interactions with other factors in the model. The subsequent divergence of the curves, for the data to the right of 100 min, is a consequence of the narrow horizontal span of any singer's data in this

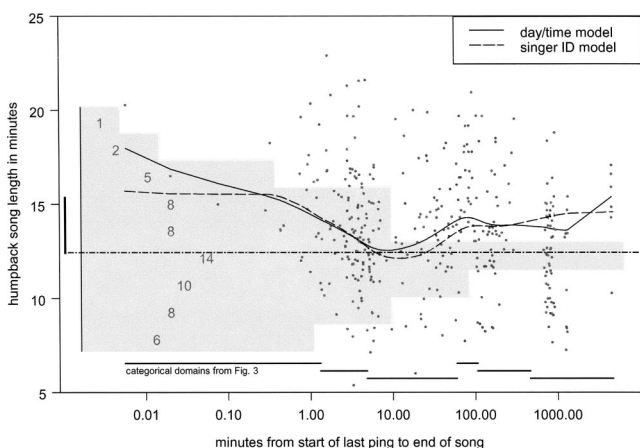


FIG. 9. The effect of time since the last ping. The solid and dashed curves represent the smoothing splines for the single-ping factor. The solid curve is from the GAM using day of year and time of day; the dashed curve is from the GAM using individual ID. The underlying histogram indicates the distribution of song lengths during morning periods before the first transmission of the day. The length of the solid bar near the left-hand axis represents the average change in the lengths of successive songs sung by a singer. The horizontal bars schematically represent the categorical groupings used in Fig. 3.

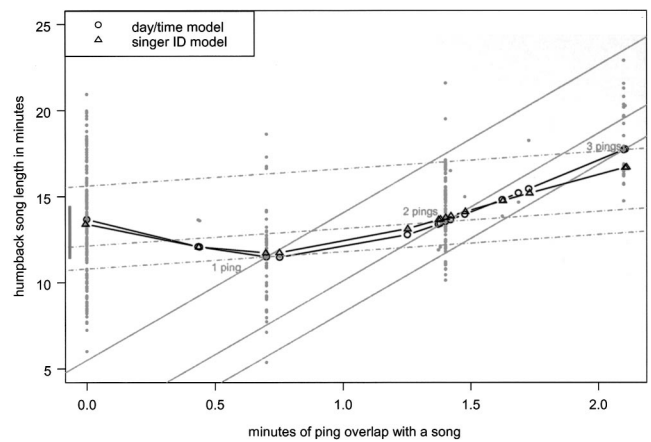


FIG. 10. The effect of ping overlap. The curves represent the smoothing splines for the minutes of ping overlap during a song. The curve marked with circles is from the GAM using day of year and time of day; the curve marked with triangles is from the GAM using individual ID. The three solid gray lines are drawn through the circles corresponding to zero, one, two, and three full pings of overlap, with a slope equal to the inverse of the duty cycle for the broadcasts ($6 \times 60 / 42$ s). This slope corresponds to the null hypothesis (singers were oblivious to the broadcasts). The three dotted lines have slopes equal to 1.0. The length of the solid bar near the left-hand axis represents the average change in the lengths of successive songs sung by a singer.

region. The singer ID factors can be arbitrarily assigned in this region, so the fitting process does not strongly constrain the shape of the time since last ping factor.

The scatterplot of song length versus the time since the last ping suggests that in addition to changes in average song length, variability of song length may increase in response to LF broadcasts. Using the categories from Fig. 3, songs ending 1.3 to 4.8 min after the last ping exhibited the largest variance in length when compared to those in the control group, songs ending more than 450 min after the last ping. Significance, however, falls short of the conventional criterion ($F_{90,66} = 1.4795$, $p = 0.0949$).

Figure 10 confirms the univariate analysis of Fig. 4: song length increases rapidly with increased ping overlap, but not quite as steeply as predicted by the oblivious hypothesis. Song length increases more rapidly than would be predicted if whales simply lengthened their songs by an amount equal to the duration of overlap. There is substantial agreement between the fitted curves from generic and singer ID models, once again indicating that this pattern is not an artifact of interactions among the modeled factors.

These data will not support models that include both the single-ping and overlap factors in a single model, because these factors covary: only two songs that ended within 12 min of the last ping were not overlapped, and only two songs that were overlapped ended more than 12 min after the last ping. However, the conclusion that ping overlap is intrinsic to the mechanism of response is contradicted by the tree-based regression of song length on time since the last ping (Fig. 3). This regression pooled songs by their mean lengths, and the 4.8-to-59-min category included songs that were and were not overlapped by pings. A t -test within this category confirmed that there was no significant difference in song length related to overlap. Thus, the single-ping model seems more broadly supported.

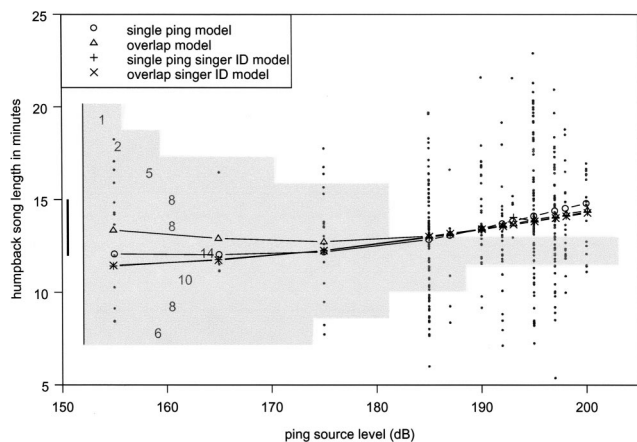


FIG. 11. The effect of ping source level. The four curves represent the smoothing splines for the source level factor from the different models. The underlying histogram indicates the distribution of song lengths during morning periods before the first transmission of the day. The length of the solid bar near the left-hand axis represents the average change in the lengths of successive songs sung by a singer.

The curves in Figs. 8 and 9 reveal the bias due to limiting the degrees of freedom of the smoothing splines used to represent the response factors. A few song length measurements were made more than a day after the last ping. This corresponds to values on the horizontal axis in excess of 2200 min. These six songs, which were recorded between 1400 and 1500 in the afternoon, had an average duration of 14.8 min. All of the songs sung 450 to 1000 min after the last ping were sung in the early morning (before 0730 h), and were characteristically shorter. Thus, the fitting algorithm used by the generalized additive model skewed the LF broadcast response factor to fit this diurnal shift in song length. This skew has two consequences: the curves do not pass through the mean of the control data, and the positive slopes of curves at the extreme right of Figs. 8 and 9 are artifacts.

The broadcast factors made stronger contributions in the single-ping and overlap models than in the ping-series model. Note, however, that the extremely strong contribution of the overlap factor (Table I, third model) must be interpreted with caution. As noted previously, the null hypothesis (no reaction to broadcasts) would predict a positive correlation between song length and minutes of overlap.

Figure 11 shows the smoothed fits of source level to song length, from four of the GAM analyses. These models yielded remarkably consistent increases in song length as source levels increased from 175 to 200 dB, the range in which most of the data were distributed. Humpback whales sang longer songs in the intervals following louder broadcasts. These models fitted the source level of the last ping to all of the data, including songs that occurred hours after the last ping. However, a regression of source level on song length for songs overlapped by a ping yielded a very similar result (slope = 0.0864, $F_{1,144} = 2.073$, $p = 0.152$).

IV. DISCUSSION

The methodological approach of this study differs in three respects from the affiliated research reported in Miller

et al. (2000). No *a priori* assumptions were made regarding the timing of responses to broadcasts. Natural factors affecting song length were explicitly modeled instead of relying on a matched sample approach to minimize their effects. Natural variation in song length was quantified, to provide a scale for assessing the magnitude of observed responses. The goal was a more comprehensive understanding of variation in song length, to provide the broadest biological context for evaluating the observed response to the LF broadcasts.

The most daunting challenge in any study of the impacts of human activities on free-ranging animals is incorporating natural variability in behavior. In the absence of adequate predictive models, this uncontrolled variation diminishes the opportunities to measure behavioral response. Thus, the statistical power of such studies depends on both the sample size and the ability to control for other factors influencing behavioral patterns. When natural factors are not taken into account, detectable responses can be obscured, and false impressions of human impact can be developed.

The matched sample approach adopted by Miller *et al.* (2000) categorized response in relation to hour-long intervals: prebroadcast, broadcast, postbroadcast. If these three intervals are identical in all other respects, then other factors affecting song length can be ignored. However, as the “matched sample” spans 3 h the assumption that all other factors remain constant is problematic. Also, it was difficult to follow free-ranging whales for 3 h, which curtailed the sample size. Sample size issues aside, the strength of the matched sample approach is also its most serious weakness: the focus on a single factor degrades the capacity of the data to provide more general insights into the natural patterns and biological significance of the behaviors under study.

The multivariate models accounted for additional factors and used songs from a much larger number of animals. However, the multivariate models lose the intrinsic control inherent in the matched sample, and are vulnerable to biases due to unmeasured factors. Factors that remain constant for 3 h but vary within the field season are no longer controlled. This concern was addressed by comparing results across several models that utilized different combinations of factors. When the fitted curves for one factor retain the same shape across models, the results are less likely to represent bias due to unmeasured factors or artifacts of the model structure.

In spite of methodological differences, the results presented here generally agree with the findings of Miller *et al.* (2000). Both studies indicate that humpback whales increase their song length in response to LF broadcasts. In this study, statistically significant differences were documented that lasted up to 2 h after the last broadcast. Source level was a significant factor in humpback responses, and higher source levels were associated with longer songs.

Additionally, this study documents that the magnitude of the response was well within the range of variation in song lengths in the absence of LF broadcasts. The responses were of the same magnitude as the average difference in the lengths of successive songs sung by an individual. The modeled responses to broadcasts also fell within the range of variation in song length observed during periods when many hours had elapsed since the last ping.

These results differ from Miller *et al.* (2000) by documenting that increases in song length were contingent on the timing of pings in relation to songs. Songs that were overlapped in the latter portion of the song showed the greatest increase in song length. Songs that were overlapped by one ping, near the start of the song, showed no increase in song length. The dependency of response to stimuli on the phase of an animal's behavior has been observed in frog playback experiments: "stimuli occurring too soon after the end of a call (inhibitory phase) are postulated to increase the delay until the next call onset, while those occurring later (in the excitatory phase) decrease the delay by stimulating a call" (Brush and Narins, 1989).

A delayed response to the LF broadcasts was also documented. The largest increases in song length were observed in songs that were sung between 1 and 2 h after the last ping. This result was based on 38 songs that were evenly distributed throughout the entire experimental period, which decreases the likelihood that it could be attributed to some other factor. It should be further noted that the three exceptionally long songs, which were excluded from the analyses, provide additional evidence for this delayed response. Two were sung 1 to 2 h after the last ping, while the third ended 37 min after the last ping. It was not possible to determine whether the delayed response scaled with the number of pings, because most ping series were of the same length.

Aside from this delayed response, other measures failed to indicate cumulative effects from the LF broadcasts. The duration of the ping series preceding songs was tested as an experimental factor. It did not provide as strong or consistent a predictor as the minutes since the last ping. This suggests that the song length response depends solely on the most recent ping, and not the immediate history leading up to that ping. The modeled seasonal and diurnal factors do not show trends that can plausibly be explained by cumulative exposure to pings. The increase in song length from early morning to afternoon was the same on days with and without pings. The seasonal trend was not unidirectional, and it appears to be correlated with local humpback population density. Finally, idiosyncratic differences among singers did not correlate with duration of potential prior exposure.

These data provide clues regarding the biological significance of song length. Humpback song length increased on days with higher local population density, and also during hours of day with higher social activity (afternoon). Similar correlations between call length and chorus density have been observed in other species, such as gray treefrogs (Welch *et al.*, 1998). This pattern could indicate a compensatory response to increased ambient noise, a competitive response to other singers' displays, or mutual correlations of chorus density and song length with a third factor, such as the availability of potential mates. Humpback responses to LF broadcasts can be viewed as consistent with these mechanisms, as if singers reacted to the pings as they would to another singer. However, better understanding of the observed responses will require more detailed studies of singing behavior and the social function of this display. The modest scale of the measured responses to LF broadcasts may reflect high fitness costs to changes in singing behavior. Male singing behavior

may appear relatively unchanged under noisome conditions because more dramatic alterations would diminish mating success.

It has been suggested previously that song length provides an index of condition because of the constraints that song structure imposes on opportunities for respiration (Chu *et al.*, 1986; Chu, 1988). However, these data show that song length is highly variable, and may play a dynamic role in social signaling. Although these analyses did reveal differences among singers in average song length, all singers exhibited substantial variation. Song length does not seem to be a rigidly stereotyped advertisement or tightly constrained by physiological condition.

The evidence of a response that scaled with source level poses questions. In the GAM models presented here, source level was assumed to have a lasting effect on song length, regardless of the time elapsed since the most recent ping. These GAM models also indicated that there was a delayed response to LF broadcasts, and these effects subsided 2 h after the last broadcast. Future research might clarify the effects of source level on the magnitude and duration of the delayed response. A more complex model would estimate the decay of source level effects as a function of the time since the last ping.

The literature on marine mammal responses to noise has not adequately addressed the interrelated effects of source level, proximity, and received level. Estimation of received level requires precise knowledge of the depths of singing whales. Position and depth of singers can be measured using acoustic localization methods. The accuracy of such methods is contingent upon developing high-resolution models for the positions of hydrophone elements in towed arrays. For these data, the hydrophones were embedded in a long cable and towed by a ship executing complex maneuvers in order to approach whales. Ongoing research effort is focused on array shape estimation, automatic detection, and localization.

Future studies should incorporate provisions to study delayed responses by varying the duration of exposure and providing for longer-term monitoring of behavioral responses. The ability of humpbacks to orient and navigate may be compromised by exposure to explosions (Todd *et al.*, 1996), even when visual observations did not detect altered residency or movement patterns in feeding areas while the whales were being exposed to the sounds (Malme *et al.*, 1985). "This suggests that caution is needed in interpreting the lack of visible reactions to sounds as an indication that whales are not affected, or harmed by an acoustic stimulus" (Todd *et al.*, 1996), and underscores the importance of examining both short-term and long-term behavioral evidence.

Rational environmental policy requires reliable measures of potential impact, combined with a plausible interpretation of their demographic significance. These results offer a detailed picture of short-term response in the context of behavioral variation observed in the absence of the stimulus. These responses were relatively brief in duration, with all observed effects occurring within 2 h of the last ping. Some changes in behavior can be expected for any perceptible stimulus, especially one associated with a large ship maneuvering nearby. The effects documented here were revealed by

careful statistical analysis, but they were not salient to the acoustic observers on the scene. Dramatic changes in humpback singing behavior would have demographic consequences, but the effects documented here do not seem to pose this risk.

ACKNOWLEDGMENTS

The successful completion of this multi-faceted research project was the result of close cooperation by a large team. The project included personnel from Cornell University, Marine Acoustics, Inc., Woods Hole Oceanographic Institution, Harvard University, the University of Hawaii, the University of Washington, the Naval Facilities Engineering Service Center, Raytheon Corp., and the US Navy.

The following research personnel represented Cornell University in the collection of the field data: Carol Carson, Christine Gabriele, David Larom, and Scott Sinclair. Lars Bejder, Nicoletta Biassoni, Frank Cipriano, Bete Jones, Patrick Miller, Amy Samuels, and Peter Tyack, represented Woods Hole Oceanographic Institution in the collection of the shore station visual data and vessel-based focal follow data. William Ellison and Stan Laback from Marine Acoustics Inc. assisted in the collection of the acoustic and environmental field data. Kathy Dunsmore, Tom Calupca, and Harold Mills developed software for acoustic data collection and analysis. We owe special thanks to Joseph Johnson, the program manager who directed and motivated the entire project. We are indebted to the Navy for access to the LFA system and use of the R/V CORY CHOUET. We gratefully acknowledge the important contributions of Chief Warrant Officer Mike Lamczyk and the Navy crew in the successful collection of visual sighting, environmental, and acoustic data. We thank the captain and crew of the R/V CORY CHOUET, including Jeff Mentzer, Mark Test, Tom Oliphant, Trish Mackay, John Demler, Bill Cermak, Mark Fuller, Howard Hanway, Roger Buckmann, Larry Lewis, Rene Osias, Charles Bain, Jim Bennet, Dave Bentley, Rick Mroch, Terry Miranda, and Keith Nicholas. We thank personnel from the State of Hawaii's Division of Land and Natural Resources, the Hawaiian Islands Humpback Whale National Marine Sanctuaries, and the National Marine Fisheries Service for advice and constructive criticism. Funding for the Cornell research was provided by the Strategic Environmental Research and Development Program (SERDP) through the Office of Naval Research Grant No. N00014-97-1-1027.

Au, W. W. L., and Green, M. (2000). "Acoustic interaction of humpback whales and whale-watching boats," *Mar. Environ. Res.* **49**, 469–481.

Baker, C. S., and Herman, L. M. (1984). "Aggressive behavior between humpback whales (*Megaptera novaeangliae*) wintering in Hawaiian waters," *Can. J. Zool.* **62**, 1922–1937.

Barlow, J. (1995). "The abundance of cetaceans in California waters: Part I. Ship surveys in summer and fall of 1991," U.S. National Marine Fisheries Service Fish. Bull. **93**, 1–14.

Biassoni, N. (2000). "Preliminary results of the effects of SURTASS-LFA sonar on singing humpback whales," W. H. O. I. Tech. Report 2000-06, Woods Hole, MA.

Brown, W. D., Wideman, J., Andrade, M. C. B., Mason, A. C., and Gwynne, D. T. (1996). "Female choice for an indicator of male size in the song of the black-horned tree cricket, *Oecanthus nigricornis* (Orthoptera: Gryllidae: Oecanthinae)," *Evolution* (Lawrence, Kans.) **50**, 2400–2411.

Brush, J. S., and Narins, P. M. (1989). "Chorus dynamics of a neotropical amphibian assemblage: comparison of computer simulation and natural behavior," *Anim. Behav.* **37**, 33–44.

Catchpole, C. K. (1980). "Sexual selection and the evolution of complex songs among European warblers of the genus *Acrocephalus*," *Behaviour* **74**, 149–166.

Cerchio, S., Jacobsen, J., and Norris, T. (2001). "Geographic and temporal variation in songs of humpback whales (*Megaptera novaeangliae*): synchronous change in Hawaiian and Mexican breeding assemblages," *Anim. Behav.* **62**, 313–329.

Chambers, J. M., and Hastie, T. J. (editors) (1991). *Statistical Models in S* (Wadsworth and Brooks, Pacific Grove, CA).

Chu, K. C. (1988). "Dive times and ventilation patterns of singing humpback whales," *Can. J. Zool.* **66**, 1322–1327.

Chu, K., and Harcourt, P. (1986). "Behavioral correlations with aberrant patterns in humpback whale *Megaptera novaeangliae* songs," *Behav. Ecol. Sociobiol.* **19**, 309–312.

Clapham, P. J. (1996). "The social and reproductive biology of humpback whales: an ecological perspective," *Mammal. Rev.* **26**, 27–49.

Clark, C. W., and Tyack, P. L. (1998). "Low-frequency sound scientific research program. Phase III: responses of humpback whales to SURTASS LFA off the Kona coast, Big Island Hawaii I 26 February–31 March 1998," Quick Look Report. Marine Acoustics, Inc., 809 Aquidneck Ave., Middletown, RI.

Corkeron, P. J. (1995). "Humpback whales (*Megaptera novaeangliae*) in Hervey Bay, Queensland-behavior and responses to whale-watching vessels," *Can. J. Zool.* **73**, 1290–1299.

Craig, A. S., and Herman, L. M. (1997). "Sex differences in site fidelity and migration of humpback whales (*Megaptera novaeangliae*) on the Hawaiian Islands," *Can. J. Zool.* **75**, 1923–1933.

Darling, J. D. (1983). "Migrations, abundance and behavior of Hawaiian humpback whales, *Megaptera novaeangliae* (Borowski)," Ph.D. Thesis, University of California, Santa Cruz.

Eiriksson, T. (1994). "Song duration and female response behavior in the grasshopper *Omocestus viridulus*," *Anim. Behav.* **47**, 707–712.

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

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

Frankel, A. S., and Clark, C. W. (2002). "ATOC and other factors affecting the distribution and abundance of humpback whales (*Megaptera novaeangliae*) off the north shore of Kauai," *Marine Mammal Sci.* **18**, 644–662.

Frankel, A. S., Clark, C. W., Herman, L. M., and Gabriele, C. M. (1995). "Spatial distribution, habitat utilization, and social interactions of humpback whales, *Megaptera novaeangliae*, off Hawaii, determined using acoustic and visual techniques," *Can. J. Zool.* **73**, 1134–1146.

Frumhoff, P. (1983). "Aberrant songs of humpback whales (*Megaptera novaeangliae*): clues to the structure of humpback songs," in *Communication and Behavior of Whales*, edited by R. Payne (Westview, Boulder, CO), pp. 81–127.

Gabriele, C. M. (1992). "The behavior and residence characteristics of reproductive classes of humpback whales (*Megaptera novaeangliae*) in the Hawaiian Islands," M.A. thesis, University of Hawaii, Honolulu.

Gentner, T. Q., and Hulse, S. H. (2000). "Female European starling preference and choice for variation in conspecific male song," *Anim. Behav.* **59**, 443–458.

Hastie, T. J., and Tibshirani, R. J. (1998). *Generalized Additive Models* (Chapman and Hall/CRC, Boca Raton, FL).

Helweg, D. A., Frankel, A. S., Mobley, J. R., Jr., and Herman, L. M. (1992). "Humpback whale song: our current understanding," in *Sensory Systems of Marine Mammals*, edited by J. Thomas, R. Kastelein, and A. Supin (Plenum, New York), pp. 459–483.

Herman, L. M., and Antinaja, R. C. (1977). "Humpback whales in the Hawaiian breeding waters: population and pod characteristics," *Sci. Rep. Whales Res. Inst. Tokyo.* **29**, 59–85.

Klump, G. M., and Gerhardt, H. C. (1987). "Use of non-arbitrary acoustic criteria in mate choice in female gray tree frogs," *Nature* (London) **326**, 286–288.

Malme, C. I., Miles, P. R., Tyack, P. L., Clark, C. W., and Bird, J. E. (1985). "Investigation of the potential effects of underwater noise from petroleum industry activities on feeding humpback whale behavior," Report to U.S.

- Department of the Interior, Minerals Management Service, Anchorage, Alaska (NTIS PB86-218385).
- The Mathworks, Inc. (1999). Matlab 5.3., Natick, MA.
- Maybaum, H. L. (1990). "Effects of a 3.3 kHz sonar system on humpback whales, *Megaptera novaeangliae*, in Hawaiian waters," EOS **71**, 92.
- Maybaum, H. L. (1993). "Responses of humpback whales to sonar sounds," J. Acoust. Soc. Am. **94**, 1848–1849.
- Miller, P. J. O., Biassoni, N., Samuels, A., and Tyack, P. L. (2000). "Humpback whales sing longer songs when exposed to LFA sonar," Nature (London) **405**, 903.
- Mobley, Jr., J. R., Bauer, G. B., and Herman, L. M. (1999). "Changes over a ten-year interval in the distribution and relative abundance of humpback whales (*Megaptera novaeangliae*) wintering in Hawaiian waters," Aquat. Mammals **25**, 63–72.
- Mobley, J. R., Jr., Forestall, P. H., and Grotefendt, R. A. (1994). "Results of 1993 aerial surveys in Hawaiian waters," University of Hawaii at West Oahu.
- Mobley, J. R., Jr., Forestall, P. H., and Grotefendt, R. A. (1995). "Preliminary results of 1993 and 1995 aerial surveys of Hawaiian waters," report to the ATOC-MMRP.
- Mobley, Jr., J. R., Herman, L. M., and Frankel, A. S. (1988). "Responses of wintering humpback whales *Megaptera novaeangliae* to playback of recordings of winter and summer vocalizations and of synthetic sound," Behav. Ecol. Sociobiol. **23**, 211–224.
- Navy, Dept. of the (2001). "Final overseas environmental impact statement and environmental impact statement for surveillance towed array sensor system low frequency active (SURTASS LFA) sonar," SURTASS LFA Sonar Program Manager, Chief of Naval Operations.
- Norris, T. F. (1994). "The effects of boat noise on the acoustic behavior of humpback whales (*Megaptera novaeangliae*)," J. Acoust. Soc. Am. **96**, 3251.
- Payne, R. S., and McVay, S. (1971). "Songs of humpback whales," Science **173**, 585–597.
- Payne, K., and Payne, R. (1985). "Large scale changes over 19 years in songs of humpback whales in Bermuda," Z. Tierpsychol. **68**, 89–114.
- Payne, K., Tyack, P. L., and Payne, R. (1983). "Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): a detailed analysis of two seasons in Hawaii," in *Communication and Behavior of Whales*, edited by R. Payne (Westview, Boulder, CO), pp. 9–57.
- Richardson, W. J., Greene, C. R., Malm, C. I., and Thompson, D. H. (1995). *Marine Mammals and Noise* (Academic, San Diego, CA).
- Statistical Sciences, Inc. (2000). "S-Plus 2000 Professional Release 1."
- Taigen, T. L., and Wells, K. D. (1985). "Energetics of vocalizations by an anuran amphibian (*Hyla versicolor*)," J. Comp. Physiol. **155**, 163–170.
- Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (2000). "Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations," J. Acoust. Soc. Am. **107**, 1286–1300.
- Todd, S., Stevick, P., Lien, J., Marques, F., and Ketten, D. (1996). "Behavioural effects of exposure to underwater explosions in humpback whales (*Megaptera novaeangliae*)," Can. J. Zool. **74**, 1661–1672.
- Tyack, P. L. (1981). "Interactions between singing Hawaiian humpback whales and conspecifics nearby," Behav. Ecol. Sociobiol. **8**, 105–116.
- Tyack, P. L. (1983). "Differential response of humpback whales, *Megaptera novaeangliae*, to playback of song or social sounds," Behav. Ecol. Sociobiol. **13**, 49–55.
- Welch, A. M., Semlitsch, R. D., and Gerhardt, H. C. (1998). "Call duration as an indicator of genetic quality in male gray tree frogs," Science **280**, 1928–1930.
- Winn, H. E., and Winn, L. K. (1978). "The song of the humpback whale (*Megaptera novaeangliae*) in the West Indies," Mar. Biol. (Berlin) **47**, 97–114.