A method for detecting whistles, moans, and other frequency contour sounds

David K. Mellinger[a)]

Cooperative Institute for Marine Resources Studies, Oregon State University and Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, 2030 Southeast Marine Science Dr., Newport, Oregon 97365

Stephen W. Martin

Space and Naval Warfare Systems Center Pacific, 53560 Hull Street, San Diego, California 92152

Ronald P. Morrissey

Naval Undersea Warfare Center, Building 1351, Newport, Rhode Island 02841

Len Thomas

Centre for Research into Ecological and Environmental Modelling, University of St. Andrews, St. Andrews, Fife KY16 9LZ, Scotland

James J. Yosco

SAIC Inc., 23 Clara Drive, Mystic, Connecticut 06355

An algorithm is presented for the detection of frequency contour sounds—whistles of dolphins and many other odontocetes, moans of baleen whales, chirps of birds, and numerous other animal and non-animal sounds. The algorithm works by tracking spectral peaks over time, grouping together peaks in successive time slices in a spectrogram if the peaks are sufficiently near in frequency and form a smooth contour over time. The algorithm has nine parameters, including the ones needed for spectrogram calculation and normalization. Finding optimal values for all of these parameters simultaneously requires a search of parameter space, and a grid search technique is described. The frequency contour detection method and parameter optimization technique are applied to the problem of detecting "boing" sounds of minke whales from near Hawaii. The test data set contained many humpback whale sounds in the frequency range of interest. Detection performance is quantified, and the method is found to work well at detecting boings, with a false-detection rate of 3% for the target missed-call rate of 25%. It has also worked well anecdotally for other marine and some terrestrial species, and could be applied to any species that produces a frequency contour, or to non-animal sounds as well. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3531926]

I. INTRODUCTION

Automatic detection of animal vocalizations is used for analyzing sound signals in real time during surveys (e.g., Leaper *et al*., 2000; Clark *et al*., 2010), for analyzing large recorded data sets (e.g., Mellinger *et al*., 2004), for real-time warning systems to mitigate the impact of human activities on wildlife (e.g., Clark *et al*., 2007), and a variety of other purposes. It is also useful for estimating the population density of a species in a given area: One method (Marques *et al*., 2009) involves counting the number of calls from the target species in a fixed period of time, then using the average call rate and the detection distances to estimate the number of individuals per unit area. Detection of transient sounds is also used or proposed in non-bioacoustic applications such as detection of illegal fishing, music analysis, and the detection and localization of gunshots.

One common type of vocalization is a frequency contour—a sound comprising one or more harmonic components that are not impulsive but instead last for some duration. Depending on their frequency, duration, and timbre characteristics, such sounds are variously described as whistles (e.g., Sayigh *et al*., 2007), squeaks (Kullenberg, 1947), trumpets (Glaeser, 2009), moans (Cummings and Thompson, 1971) and so on. The appearance of such sounds in a spectrogram is a frequency peak which varies continuously with time, forming a contour. Automatic detection of such sounds could potentially have applications in the areas mentioned above, and because frequency contours are found so commonly among animal species, such a detector could potentially have wide application.

Several methods have been proposed to detect such sounds. Matched filtering is optimal when the sound signal is invariant and is embedded in white Gaussian noise. In most bioacoustic applications, however, neither of these conditions applies. Another template-matching method is spectrogram correlation (Mellinger and Clark, 2000), which allows for some variability in the frequency contour to be detected but

---

[a)]Author to whom correspondence should be addressed. Electronic mail: David.Mellinger@oregonstate.edu

is still best employed to detect relatively stereotyped call types. A more general technique for finding frequency contours is edge detection (e.g., Gillespie, 2004), which uses techniques derived from image processing to find the edges of a frequency contour in a spectrogram image. Once the contour has been extracted from the spectrogram, classification techniques are applied to determine whether the frequency contour sound is of the desired type. Buck and Tyack (1993) describe a contour extraction algorithm that extracts the highest point in the spectrogram in each time frame, then checks whether the detected peak is the first or second harmonic. Mallawaarachchi *et al.* (2008) developed a contour-tracking system using a Kalman filter that tracks the frequency, frequency change, and frequency acceleration of a whistle in a spectrogram. Madhusudhana *et al.* (2008) describe a two-stage algorithm using forward and backward contour tracking to find the best set of contours to fit sets of spectral peaks.

Here we describe two advances. First we describe an algorithm that extends that of Buck and Tyack (1993) for detecting frequency contours in a sound signal. This algorithm employs several parameters that control its operation—parameters that can be difficult to choose values for because each can interact with others in determining the operating characteristics of the algorithm. Because of this difficulty, we next describe an iterative method for choosing parameter values that results in an optimal set of values for a given detection task. As a test case, the detection algorithm and parameter optimization procedure are applied to the problem of detecting minke whale (*Balaenoptera acutorostrata*) boing vocalizations off Hawaii.

II. FREQUENCY CONTOUR DETECTION ALGORITHM

The goal is to find frequency contours in the spectrogram that are within a given frequency range $[f_0, f_1]$ starting at a given time t_0. A spectrogram $S(t,f)$ (Oppenheim and Shafer, 1975) is calculated using a Hamming window and frame size s (in seconds) and overlap of half the frame size. Spectrum-flattening normalization (Mellinger, 2004) with a time constant t_{norm} is applied to each frequency band of the spectrogram to obtain a noise-flattened spectrogram $\hat{S}(t,f)$. Briefly, this works by computing, in each frequency bin, a running average over successive spectrogram slices, then subtracting this average from the spectrogram to arrive at $\hat{S}(t,f)$.

For detecting frequency contours at or after time t_0, the spectrum (i.e., normalized spectrogram slice) at t_0 is examined to find all local maxima between f_0 and f_1. These peaks form the set of *candidate contours*. Each peak is then checked to ensure that it is a certain level h above the background noise, where the background noise is the median of all normalized spectrum values in the time slice. Normalization of the spectrogram over a relatively long time period (say, $>3t_{norm}$) as described above essentially flattens, or whitens, the spectrum, so that the median spectrum value is likely to represent the level of noise or impulsive noise sounds rather than any spectrum-altering filter or other artifact in the sound acquisition process. Actually the noise estimate need not come from the median of the spectrum, which is the 50th-percentile value; it can be any percentile p. Each peak is
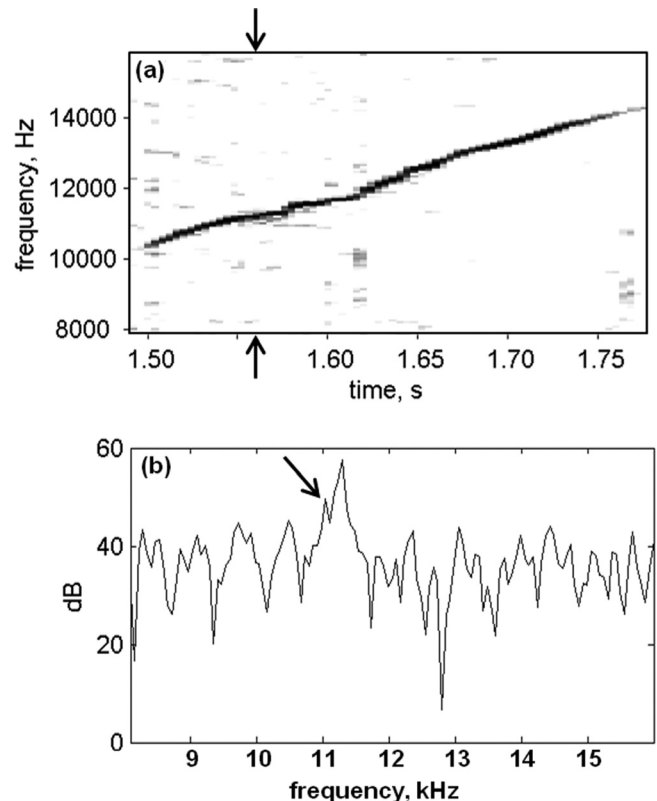

FIG. 1. A subsidiary spectral peak. (a) Spectrogram of a dolphin whistle, probably from a spotted dolphin (*Stenella* sp.). Arrows indicate the time slice of the spectrum in (b). Spectrogram parameters: sample rate 32 kHz, frame size and FFT size 512 samples, 75% overlap, Hamming window, for a filter bandwidth of 254 Hz. (b) Spectrum of a single time slice in the frequency band near the dolphin whistle. Arrow indicates a local maximum in the spectrum that is 10 dB or more above the median background level, but that is a subsidiary of a larger peak and therefore should not be considered as a separate candidate frequency contour.

checked to determine whether it is smaller than other spectrum values within a neighborhood N (in hertz) in the spectrum, discarding it from the set of candidate contours if so. This prevents small bumps on the side of a large peak from being considered independent candidate contours (Fig. 1). This process narrows the set of candidate contours.

The processes are repeated in successive time slices (successive spectra). The peaks from a given spectrum are matched to contours from previous spectra by fitting a curve through the peaks comprising a given contour and extrapolating to the succeeding time slice. Any mathematical function such as a polynomial or a spline may be used in this fit; a line (first-order polynomial) is used in the test case below. After fitting, extrapolated values that are within some distance f_d of the existing candidate contour are accepted as part of the contour, while peaks that do not match an existing contour become seeds for new contours. Using an extrapolated frequency in this manner allows the algorithm to handle crossing frequency contours from multiple animals, keeping each contour on its proper trajectory after the cross. It also facilitates bridging short gaps in the sound as described below.

A candidate contour is thus extended through successive time slices of the normalized spectrogram by (a) finding new peaks, (b) matching the frequency of a peak with the expected frequency of the contour, and (c) calculating a new expected

frequency of the contour. When step (b) fails and there is no peak close to the expected frequency, then one of two approaches may be taken. The first approach is to terminate the contour at this time. This approach is simple and works well enough when it is not important whether contours with gaps are detected as a single entity or as multiple successive contours. The second approach is to wait for some short period of time t_{gap}, calculating the expected frequency of a candidate contour at each step, and see if the candidate contour continues at some later time slice in the spectrogram. If so, the candidate contour may simply be extended over the silent period and continued. With this approach, the candidate contour is considered to have ended when there are no spectral peaks that match the contour's expected frequency for more than t_{gap} seconds. This second approach is useful when contours might have gaps in time, as often happens with dolphin whistles, but one wishes to continue the contour over such gaps provided they are not too long (i.e., not longer than t_{gap}).

When a candidate contour terminates, it is registered as a *detected contour* (no longer just a "candidate" one) if it has persisted for a certain minimum duration d. This criterion prevents the numerous small blips of noise typically seen in spectrograms, including normalized spectrograms, from being registered as large numbers of very short-duration contours.

Candidate contours also must persist for a minimum time t_{sep}, to count as independent contours. This is enforced because a contour often will have secondary "contours" associated with it that are no more than small bits of noise in the normalized spectrogram on the side of the main contour (Fig. 2), and these bits of noise should not be registered as contours in their own right. To some extent this need is handled by the exclusion of subsidiary peaks described above, but it was found that further exclusion was necessary to prevent registering spurious contours.

Figure 3 shows the results of running this detection algorithm on an example sound consisting of three harmonics of a dolphin whistle. The algorithm correctly finds most of the frequency contours present and ignores the clicks (impulsive sounds) that are also present. It fails, however, to detect the faint contours present in the early parts (<2.3 s) of the third

FIG. 3. Results of frequency contour detection. (a) Spectrogram of whistles and clicks of a Risso's dolphin (*Grampus griseus*). The whistle has three harmonics. Spectrogram parameters same as in Fig. 2. (b) Output of the frequency contour detection algorithm for this sound (with spectrogram still apparent as faint image in background), with black lines indicating the detected contours. Most extant contours were detected, but the continuation of the second harmonic around 15 kHz after 2.2 s was missed.

harmonic and the late parts (>2.2 s) of the second harmonic. In addition, the first harmonic is detected as two separate sections due to sudden jump in frequency of the frequency contour at about 2.08 s.

Figure 4 shows the result when the algorithm was applied to a recording containing two crossing frequency contours. Because of its tracking of frequency trajectories and maintenance of multiple candidate frequency contours, the algorithm was able to follow the descending contour through the intersection separately from the ascending contour. There were three time steps (three time slices) in which the peak
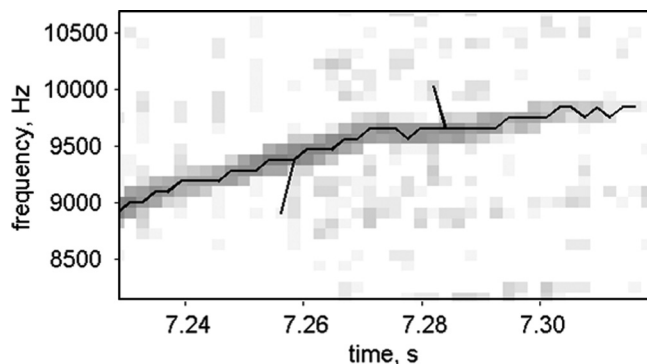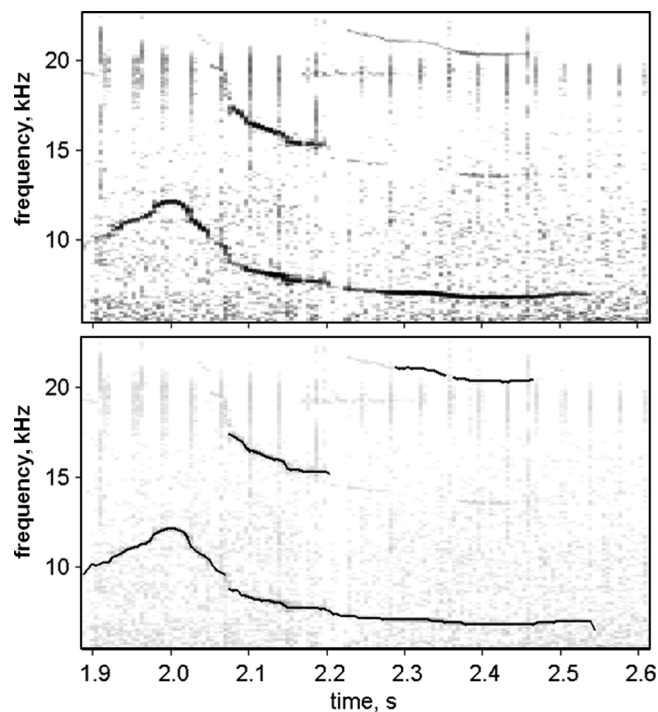

FIG. 2. Spurious detection of contours. This shows a spectrogram of a whistle from a Risso's dolphin (*Grampus griseus*) together with the frequency contours found by this algorithm when there is no restriction on the length of time or distance in frequency that a contour must be independent from another. There are two spurious contours detected, one at 7.26 s and one at 7.28 s. Spectrogram parameters same as in Fig. 1, except the sample rate is 48 kHz and the filter bandwidth is 381 Hz.
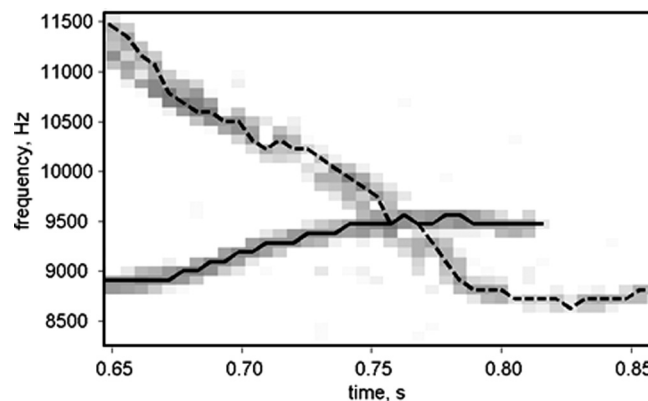

FIG. 4. Crossing contours. This shows a spectrogram of whistles from two Pacific white-sided dolphins (*Lagenorhynchus obliquidens*) that cross. The detection algorithm correctly tracks each whistle across the point where they intersect and continues each one after the cross. Spectrogram parameters same as in Fig. 2.

TABLE I. Parameters used to calculate the spectrogram, normalize it, and apply the frequency contour detection algorithm.

Parameter	Description
s	Spectrogram frame size (s)
t_{norm}	Time constant for normalization (s)
f_0, f_1	Frequency bounds (Hz)
p	Percentile for estimating background noise (0-100)
h	Level above that estimate (dB)
N	Neighborhood width (Hz)
f_d	Frequency difference from one step to the next (Hz)
d	Minimum duration (s)
k	Duration for estimating next spectral peak location (s)
t_{sep}	Minimum time that contours must be separate from other contours (s)

spectrum value was included in both contours, after which the two contours split apart again.

III. PARAMETER OPTIMIZATION PROCEDURE

The detection algorithm described above contains a large number of parameters (Table I). Because these parameters interact in complex ways, it can be difficult to find optimum values for all of them simultaneously for calls of a target species. For instance, the spectrogram frame size s interacts with the minimum contour duration d, particularly when the duration of each spectrogram frame is a large fraction of d. Similarly, the percentile for estimating noise p interacts with the minimum level h above that estimate: the larger p is, the more noise is included in the estimate, and the lower h needs to be for detection of a given contour.

Because of these many mutually interacting parameters, finding a set of optimum parameter values is not straightforward. During initial development and testing of the algorithm, the authors chose values manually, using heuristics to guide the detection. This caused them to realize the complexity of choosing the correct parameters and led to devising and employing an automated search.

The automated search operates as follows. For each parameter, a small (<10 element) set of *allowable values* is defined based on the range of values that worked well in preliminary manual testing, plus an additional value just higher and just lower than this range. The values were spaced roughly logarithmically through the range. Search starts with a random choice of a parameter set from among its allowable values (i.e., one random value for each parameter from among that parameter's allowable values). Performance is evaluated for this parameter set as described below; performance is also evaluated at adjacent neighbors in the parameter space—that is, for each parameter, the next higher and next lower allowable value is tried, with other parameters remaining fixed. The search moves to the neighboring parameter set with the best performance. This process of evaluating performance at each neighbor in the parameter space and moving to the best-performing one is iterated until performance stops improving. This iterative loop—starting with a random parameter set and repeatedly moving to the best-performing neighbor until performance stops improving—is termed an *optimization cycle*. Optimization cycles are performed for a fixed number of

times based on the amount of time available for processing, and the *optimal parameter set* is the one having the best performance among all of the cycles. This search finds a locally optimal solution (a solution better than any adjacent ones in parameter space) but not necessarily the globally optimum one (the one that is best among all possible parameter sets).

Performance is measured by varying the detection threshold. Each threshold value results in a certain fraction of wrong detections (false positives) and missed calls (false negatives). As the threshold is raised, the number of wrong detections declines but the number of missed calls increases, and vice versa for lowering the threshold. The automated search described above requires that the performance of different parameter sets be compared, and this is most easily done by having a single number that acts as a metric of performance for each parameter set. Accordingly, performance was measured here by finding the threshold that resulted in 25% missed calls (false negatives) and determining the corresponding fraction of wrong detections (false positives); lower values correspond to higher performance. This fraction was the performance metric used in the parameter optimization procedure for the case study.

IV. CASE STUDY: DETECTING MINKE WHALE BOING SOUNDS

Minke whales in the central Pacific produce a vocalization known as the boing sound (Thompson and Friedl, 1982) which is complex in nature and includes frequencies from approximately 1 kHz to above 10 kHz (Rankin and Barlow, 2005). This vocalization has durations ranging from under 1 s to over 4 s and consists of an initial burst, followed by a frequency modulation component transitioning into a relatively long amplitude-modulated constant frequency. Central Pacific (Hawaiian) boings have dominant energy between 1.3 and 1.5 kHz [Fig. 5(a); Thompson and Friedl, 1982], providing initial frequency bounds for the detector. If boings were the only frequency contour in their frequency range, detecting them would
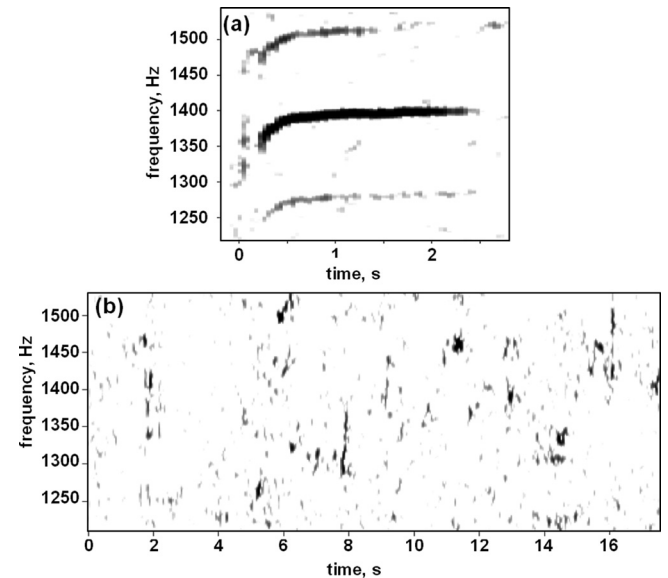
FIG. 5. Example vocalizations from northwest of Kaua'i. (a) Minke whale boing vocalization. (b) Humpback whale song. Spectrogram parameters: frame size 0.17 s, FFT size 0.68 s, overlap 75%, Hamming window, for a filter bandwidth of 24 Hz.

4058   J. Acoust. Soc. Am., Vol. 129, No. 6, June 2011

Mellinger *et al.*: Frequency contour detection method

be relatively simple, but they are not: Humpback whales (*Megaptera novaeangliae*), which are present in many of the same areas at the same times of year as minke whales, produce a wide variety of sound units as part of their song, including some in the frequency band of minke whale boings [Fig 5(b)].

The frequency contour detection algorithm and associated parameter optimization procedure were trained and tested using a group of recordings containing boings from northwest of the island of Kaua'i, Hawaii. The recordings were made between the months of February and May in 2005–2007 and comprised of eight 10-min recordings from each of 24 or 31 hydrophones, for a total of 2060 min of sound. These periods of time includes the peak boing season which is typically from February through March. The recordings analyzed were stored at a sample rate of either 6 or 12 kHz with 16-bit resolution. An analyst (J.Y.) aurally and visually (i.e., in spectrograms) reviewed the recordings and marked the time at which each boing occurred. A total of 1084 boings were found via this manual search. A fraction of these boings were duplicates, in that they consisted of one boing captured by multiple hydrophones. However, hydrophone spacing was large enough (tens of kilometers) that a given boing was significantly different upon reception at several hydrophones due to different multipath transmission, different intensities, and different background noise. For this reason, bias due to replication of inputs was considered non-significant. Interfering sounds were notably different between the hydrophones: Hydrophones closer to Kaua'i had large numbers of humpback whale vocalizations [e.g., Fig. 5(b)], while those farther away had few or none. As the hydrophones were not calibrated, background noise levels on the recordings were not known.

The parameter optimization procedure was run on the frequency contour detection algorithm; a detection by the algorithm was counted as correct if it was within 4.0 s of the time marked in the manual analysis. This time difference was necessary because the time indicated both by the algorithm and by the manual analysis was sometimes at the start of a boing, sometimes near the end, or (most commonly) somewhere in between. Boing sounds were rare enough in the recordings that the 4-s comparison window resulted in few or no spurious mis-identifications.

The optimization procedure resulted in an optimal parameter set, which was then tested using a separate set of recordings made at different times, but in the same periods of time (February to May of 2005–2007), from the same hydrophones northwest of Kaua'i. This data set comprised 38 ten-minute recordings with a total of 137 manually detected boings. This data set also contained many interfering sounds, principally from humpback whales, that were in the frequency range of minke whale boings and that sometimes triggered false detections.

V. RESULTS

The parameter optimization procedure was run for one week on a 2 GHz Intel Pentium processor, a total of 37 optimization cycles of the procedure. The optimal parameter set found by this procedure is shown in Table II.

Four of the parameters (s, h, d, and k) had a value at the boundary of their range of possible values. To ensure that

TABLE II. The range of allowable values for each parameter in the optimization procedure, and the resulting optimal parameter set after running that procedure for 37 optimization cycles.

Parameter	Range of allowable values	Optimal value found
s	0.17–0.68 s	0.17 s
t_{norm}	7–15 s	10 s
f_0	1325–1375 Hz	1350 Hz
f_1	1430–1460 Hz	1440 Hz
p	65–90	75
h	0–2 dB	2 dB
N	2–10 Hz	3 Hz
f_d	10–30 Hz	20 Hz
d	0.7–2 s	0.7 s
k	0.2–0.4 s	0.4 s
t_{sep}	0.05–0.15 s	0.10 s

the optimal value was not outside this range, each parameter was set to the next incremental value beyond the boundary value in Table II, and the detector's performance was measured. In none of these cases did the value from beyond the boundary perform better than the value shown in Table II.

This parameter set was then used for detecting minke vocalizations in the test data set containing 137 boings. Changing the threshold changed the fraction of detected boings and false detections as shown in the detection error tradeoff (DET) curve shown in Fig. 6. [The DET curve (Martin *et al.*, 1997) is like the more familiar receiver operating characteristic (ROC) curve but has error rates on both axes as well as logarithmically scaled axes.) The rate of false detections in this graph was somewhat arbitrarily scaled such that the maximum value was 100%—i.e., the number of false detections for each threshold was divided by the number of false detections, 740, that occurred with the minimum threshold used. For reference, this maximum rate corresponds to one false detection every 16.2 s. At the threshold with a 25% missed-call rate, the corresponding false-detection rate is one for every 9 min of data. In all cases for which a cause could be determined, false detections were caused by humpback whale sounds, though the sound(s) that triggered some false detections could not be identified with certainty.

VI. DISCUSSION

The detection method worked well at tracking frequency contours in examples of dolphin whistles (Fig. 3). This is
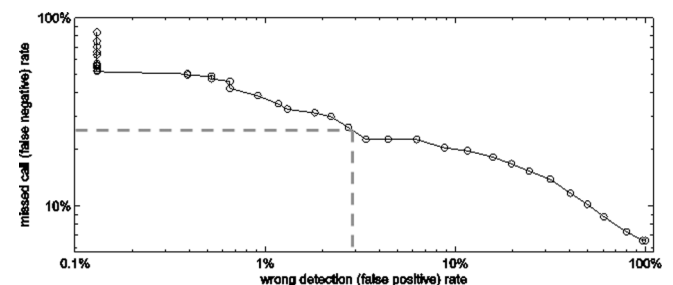

FIG. 6. The DET curve for the optimal parameter set when run on the test data set containing 137 manually identified calls. Each small circle corresponds to one threshold value. The dashed line indicates the wrong-detection rate of approximately 3% for the selected rate of 25% missed calls that was used in the optimization procedure.

J. Acoust. Soc. Am., Vol. 129, No. 6, June 2011

Mellinger *et al.*: Frequency contour detection method   4059

because the algorithm keeps track of multiple contours simultaneously—in essence, preserving multiple hypotheses about what contours exist at any instant—allowing it to track contours that cross (e.g., Fig. 4).

The algorithm also worked for the detection of minke boing sounds. Figure 6 shows that, at the chosen missed-call rate of 25% for the boing sound, the optimal parameter set resulted in a false-detection rate of approximately 3%. The false-detection rate depends on the nature of the test data set; this set of test data had good representation of chorusing humpback whales song units which contain significant energy in the selected detection frequency band and have a duration on the order of seconds. If one were to use, for example, either Gaussian white noise or summertime deep ocean sounds when very little energy resides in the selected detection band, the number of false positives would likely be greatly reduced.

The performance displayed in Fig. 6 shows that this method would be useful for a number of applications in marine mammal bioacoustics. For a data set containing numerous humpback whale song units, the false alarm rate is only 3% at a missed-call rate of 25%. This makes the method feasible for application to density estimation, depending on the density estimation method used (some do not require a low false alarm rate; Marques *et al.*, 2009). Indeed, as part of a project to estimate the density of minke whales, it is currently being used to analyze a large body of recordings from Hawaii to find the thousands of minke whale boing sounds present. It could be applied to estimate the density of any other species that produce frequency-contour calls as well.

The method described here not only detects frequency contours but also tracks their frequency over time. This information is not used in the case study here but is produced as a side-effect of the algorithm and could be used for studies of vocal behavior, communication, caller identity, or other subjects. The algorithm also uses the level of each frequency contour above background noise, which is essentially a measure of the signal-to-noise ratio (SNR) of the contour. This too could be used for other studies, such as methods for estimating detection range of vocalizing animals, methods for distinguishing successive calls of individuals over short time scales, etc.

This method could produce these types of data—detected contours and the frequencies and SNRs of those contours as a function of time—for any animal that produces frequency-contour calls. Among marine mammals this includes most delphinids, some beaked whales, some phocid seals, and other species as well. The method could also be applied to the great many terrestrial species of birds and mammals that make calls with frequency-contour components, as well as to non-animal sounds.

VII. CONCLUSION

We have demonstrated an algorithm for the detection of frequency contour sounds, as well as a method for optimizing the several parameters used in this algorithm. Subjectively, the method appears to work well for detection of dolphins whistles, as illustrated by Figs. 3 and 4. The method was quantitatively tested on a data set comprising minke whale boing sounds from the central North Pacific and was shown (Fig. 6) to have good performance in the presence of many interfering sounds from humpback whales.

To make it available in a user-friendly form, the frequency contour detection method (though not the parameter optimization procedure) has been implemented in the free software package ISHMAEL (Mellinger, 2001), version 2.0. It is hoped that this method can be usefully applied to other tasks in animal call detection and indeed detection of frequency contours of non-animal sounds as well.

Buck, J. R., and Tyack, P. L. (**1993**). "A quantitative measure of similarity for Tursiops truncatus signature whistles," J. Acoust. Soc. Am. **94**, 2497–2506.

Clark, C. W., Brown, M., and Corkeron, P. (**2010**). "Visual and acoustic surveys for North Atlantic right whales, *Eubalaena glacialis*, in Cape Cod Bay, Massachusetts, 2001–2005: Management implications," Marine Mammal Sci. **26**, 837–854.

Clark, C. W., Gillespie, D., Nowacek, D. P., and Parks, S. E. (**2007**). "Listening to their world: Acoustics for monitoring and protecting right whales in an urbanized world," in *The Urban Whale*, edited by S. D. Kraus and R. M. Rolland (Harvard University Press, Cambridge), pp. 333–357.

Cummings, W. C., and Thompson, P. O. (**1971**). "Underwater sounds from the blue whale, *Balaenoptera musculus*," J. Acoust. Soc. Am. **50**, 1193–1198.

Gillespie, D. (**2004**). "Detection and classification of right whale calls using an edge detector operating on a smoothed spectrogram," Can. Acoust. **32**, 39–47.

Glaeser, S. S. (**2009**). "Analysis and classification of sounds produced by Asian elephants (*Elephas maximus*)," Masters thesis, Portland State University, Portland, Oregon.

Kullenberg, B. (**1947**). "Sound emitted by dolphins," Nature **160**, 648.

Leaper, R., Gillespie, D., and Papastavrou, V. (**2000**). "Results of passive acoustic surveys for odontocetes in the Southern Ocean," J. Cetacean Res. Manage. **2**, 187–196.

Madhusudhana, S. K., Oleson, E. M., Soldevilla, M. S., Roch, M. A., and Hildebrand, J. A. (**2008**). "Frequency based algorithm for robust contour extraction of blue whale B and D calls," *Proceedings of the IEEE Oceans*, Kobe, Japan, pp. 8.

Mallawaarachchi, A., Onga, S. H., Chitre, M., and Taylor, E. (**2008**). "Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles," J. Acoust. Soc. Am. **124**, 1159–1170.

Marques, T. A., Thomas, L., Ward, J., DiMarzio, N., and Tyack, P. L. (**2009**). "Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville's beaked whales," J. Acoust. Soc. Am. **125**, 1982–1994.

Martin, A., Doddington, G., Kamm, T., Ordowski, M., and Przybocki, M. (**1997**). "The DET curve in assessment of detection task performance," in Eurospeech'97, Rhodes, Greece, pp. 1895–1898.

Mellinger, D. K. (**2001**). ISHMAEL 1.0 User's Guide. National Oceanographic Atmospheric Administration Technical Memorandum OAR–PMEL–120 (NOAA PMEL, Seattle), 30 p.

Mellinger, D. K. (**2004**). "A comparison of methods for detecting right whale calls," Can. Acoust. **32**, 55–65.

Mellinger, D. K., and Clark, C. W. (**2000**). "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529.

Mellinger, D. K., Stafford, K. M., Moore, S.E., Munger, L., and Fox, C. G. (**2004**). "Detection of North Pacific right whale (*Eubalaena japonica*) calls in the Gulf of Alaska," Marine Mammal. Sci. **20**, 872–879.

Oppenheim, A. V., and Schafer, R.W. (**1975**). "Power spectrum estimation," in *Digital Signal Processing* (Prentice-Hall: Upper Saddle River, New Jersey), pp. 532–576.

Rankin, S., and Barlow, J. (**2005**). "Source of North Pacific boing sound attributed to minke whales," J. Acoust. Soc. Am. **118**, 3346–3351.

Sayigh, L. S., Esch, H. C., Wells, R. S., and Janik, V. M. (**2007**). "Facts about signature whistles of bottlenose dolphins, *Tursiops truncatus*. Anim. Behav. **74**, 1631–1642.

Thompson, P. O., and Friedl, W. A. (**1982**). "A long term study of low frequency sounds from several species of whales off Oahu, Hawaii," Cetology **45**, 1–19.