DETECTION AND RECOGNITION OF NORTH ATLANTIC RIGHT WHALE CONTACT CALLS IN THE PRESENCE OF AMBIENT NOISE

Ildar R. Urazghildiiev, Christopher W. Clark and Timothy P. Krein
Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850

ABSTRACT

The problem of detection and recognition of contact calls produced by North Atlantic right whales, *Eubalaena glacialis*, is considered. A proposed solution is based on a multiple-stage hypothesis-testing technique involving a spectrogram-based detector, spectrogram testing, and feature vector testing algorithms. Results show that the proposed technique is able to detect over 80% of the contact calls detected by a human operator and to produce about 26 false alarms per 24 h of observation.

SOMMAIRE

Un problème de détection et reconnaissance des baleines noires, *Eubaleana glacialis*, en présence de bruit ambiant est étudié. Une solution proposée est basée sur une technique de test d'hypothèses en plusieurs étapes, impliquant le détecteur, des tests de spectrogramme et des algorithmes testant des vecteurs de traits. Les résultats des tests montrent que la technique proposée est capable de détecter plus de 80% des appels de contact détectés par les opérateurs humains et de produire environ 26 fausses alarmes par 24 h d'observation.

1. INTRODUCTION

Continuous monitoring of North Atlantic right whales (NARW) presence in large areas can be accomplished by passive acoustical methods using data recordings obtained from distributed autonomous hydrophone systems [1-4]. Such systems yield enormous data sets totaling many years of potential listening time, presenting an analytical challenge. Using human operators to visually and aurally evaluate data spectrograms is impractical in projects that collect huge amounts of data. Apart from this, the human operators often provide subjective and inaccurate estimates [5] so the design of effective, automated detection techniques is of critical importance.

To reduce subjectivity and to decrease the labor costs, various NARW detection methods known from the literature can be used (see e.g., [6-11]). These methods can potentially improve the detection efficiency by rejecting a huge portion of the data that contains no signal. However, as test results demonstrate, known methods do not provide the required trade-off between the probabilities of detection and false alarm. In particular, for the detection probability of 0.8, the lowest level of false alarm probability provided by the spectrogram-based detector is from 10^{-2} to 10^{-3}, depending on the impulsive noise rate [11]. For the NARW contact calls with the typical duration of 1 s, the range of probability of false alarm corresponds to 100–1000 false detections per 24 h of observation. Since all the detection events should be evaluated by a human operator, the labor costs are significant.

The goal of the research presented in this paper is to reduce the probability of false alarm in spectrogram-based detectors without negatively affecting the detection probability. The proposed technique is reduced to a multiple-stage hypotheses-testing process. In the initial stage, the spectrogram-based detector [11] is applied. The data segments accepted as signals in the initial stage are recognized using the proposed recognition technique. The hypothesis that the detected segment belongs to the known types of impulsive noise is tested in the second stage. If this hypothesis is rejected, a feature vector (FV) is extracted and tested in the final stage. Test results obtained using real data recordings are presented.

2. DATA MODEL AND PROBLEM FORMULATION

We use the data model similar to that considered in [11]. The NARW contact calls are modeled as polynomial-phase signals (PPS). Ambient noise is represented as a Gaussian process contaminated by unknown impulsive processes. A typical spectrogram of the input data containing a NARW contact call, background noise, impulsive noise and self-noise is shown in Fig. 1 (top frame).

We assume that the spectrogram-based detector is applied to the input data in the initial stage. For each 1 s data segment

$\boldsymbol{x}(t)$, the detector tests the hypotheses H_0 ("ambient noise is present") and H ("signal and ambient noise are present"). The decision-making process is reduced to computing the statistic $z(t) = z(\boldsymbol{x}(t))$ as a function of the tested data segment $\boldsymbol{x}(t)$. The statistic $z(t)$ is compared with a threshold C_D and the hypothesis H is accepted if $z(t) \geq C_D$. The detector consists of a bank of P linear 2-dimensional (2D) FIR filters with frequency responses specified by the frequency modulation of NARW contact calls [11]. The frequency responses of the FIR filters maximizing the statistics $z(t)$ are shown in Fig 1 (top frame) by the red

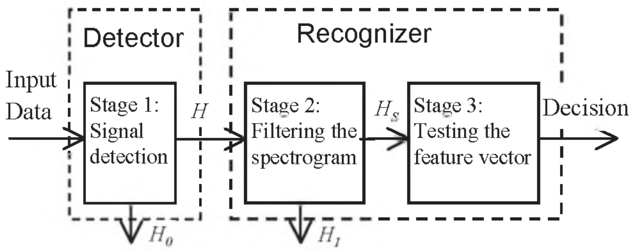

Fig. 2. A Block diagram of the proposed technique.

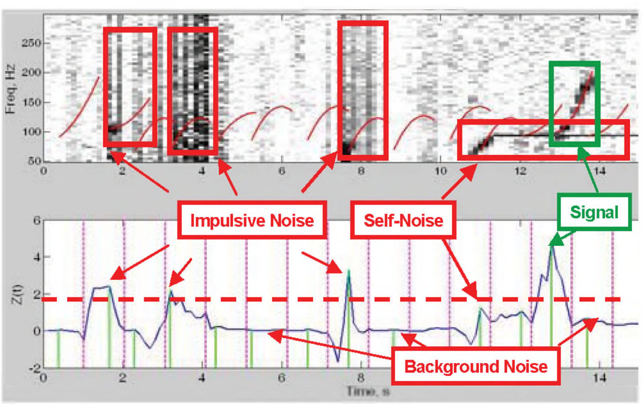

Fig. 1. A spectrogram of the input data (top frame) and the values of the statistic calculated by the spectrogram-based detector (bottom frame).

lines. The values $z(t)$ calculated from the detector output are represented in Fig. 1 (bottom frame) by green lines.

The threshold is determined by applying an optimality criterion, which is introduced based on the management goals of the detector as well as on *a priori* information. We use the Neyman-Pearson criterion, which makes it possible to minimize a false alarm probability for a given probability of detection. In practice, the threshold is set up to automatically detect 80% or more of the NARW contact calls visually detected by the human operators. The problem of choosing the threshold is beyond the scope of this paper. Instead, we focus our attention on the problem of optimizing the structure of the recognizer used in the following stages.

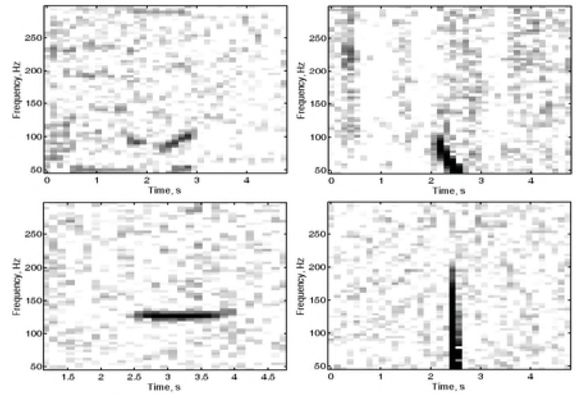

Fig. 3. The spectrograms of the typical impulsive processes: upsweep (top left), downsweep (top right), harmonic (bottom left) and wideband transient impulse (bottom right).

The data segments for which the hypothesis H_0 is accepted do not require any actions. If for a given $\boldsymbol{x}(t)$ the hypothesis H is accepted, this segment is recognized in the next stages. Only these data segments are considered hereafter (For the sake of simplicity, the time index associated with those segments is omitted). Due to the presence of impulsive noise, many segments detected in the first stage may contain no signals except noise transients. As a result, the following hypothesis can be introduced:

$$H_S : \boldsymbol{X} = \boldsymbol{S} + \boldsymbol{W}\,,\; H_I : \boldsymbol{X} = \boldsymbol{Q} + \boldsymbol{W} \qquad (1)$$

where $\boldsymbol{X}$, $\boldsymbol{S}$, $\boldsymbol{Q}$ and $\boldsymbol{W}$ are the matrixes representing the spectrograms of the data segment $\boldsymbol{x}(t)$, signal, impulsive noise, and background noise, correspondingly.

The problem can be formulated as follows: using $\boldsymbol{X}$, accept or reject the hypothesis H_S. We propose a solution based on a two-stage recognition technique. In the first stage, the hypothesis H_I is divided into the M sub-hypothesis $H_{Im}, m = 1,\ldots,M$. For each H_{Im}, a parametric model of noise is used. The models are based on the spectral properties of typical kinds of impulsive noise observed in the empirical data. Based on that model, a spectrogram-based algorithm that tests the hypothesis H_S against H_{Im} is designed. If the hypothesis H_S is accepted, the corresponding data segment is tested in the final stage. In this stage, a feature vector testing algorithm is applied. A block diagram of the proposed technique is shown in Fig. 2. The signal recognition algorithms are designed in Section 3.

3. SIGNAL RECOGNITION

Since typical noise conditions can differ for different locations, we restrict our investigations to the data

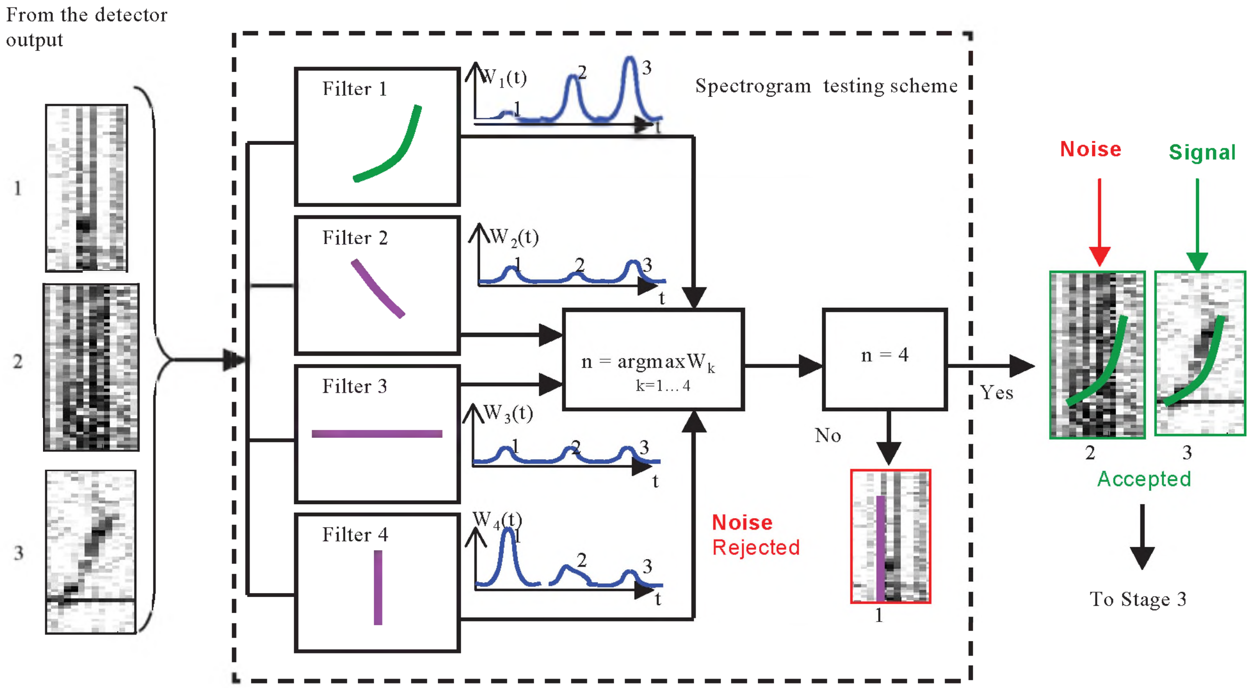

Fig. 4. Spectrogram testing scheme.

recordings collected at Cape Cod Bay, a primary habitat area for NARW. Data analysis shows that many noise transients have spectrogram-based images that are similar to the locally narrowband down-sweep and harmonic impulses as well as to wideband transients; see Fig. 3. Based on these observations, we introduce the following classes of impulsive processes: G_1 - upsweep transient, G_2 - downsweep transient, G_3 - constant frequency tone transient and G_4 - wideband transient. The class G_1 represents the signals and the classes G_2 - G_4 represent impulsive noise.

To design a spectrogram-testing algorithm, we use a strategy similar to the generalized likelihood ratio test. For each class G_m, the statistic $w(m,\boldsymbol{X})$, m = 1,…,4 is computed. If $w(1,\boldsymbol{X}) < w(k,\boldsymbol{X})$, k = 2,3,4, then the hypothesis H_I is accepted and the testing procedure is terminated for a given data segment. Otherwise the data segment is tested in the final stage.

The spectrogram testing scheme is shown in Fig. 3. It is similar to the spectrogram-based detector in the sense that it consists of a bank of 4 linear 2D FIR filters. The impulsive responses of the filters are specified by the phase structure of the typical impulsive processes, Fig. 3.

A FV is tested in the final stage. Features being used should contribute most to discrimination between signals and noise and should be easy to extract. Because of the lack of *a priori* information regarding variability of NARW contact calls, feature selection is a difficult problem. In this paper, we use features similar to those used by human operators when visually analyzing the spectrogram. Let the symbol $\boldsymbol{v} = (v_1,\ldots,v_K)^T$ denote the K-dimensional feature vector. We introduce the feature space with the dimension of K = 11 and with the features represented in Table I. The spectrogram of a NARW contact call and some features extracted from the spectrogram are also displayed in Fig. 5.

To design the FV recognition algorithm, the following approach has been used. For most of the NARW contact

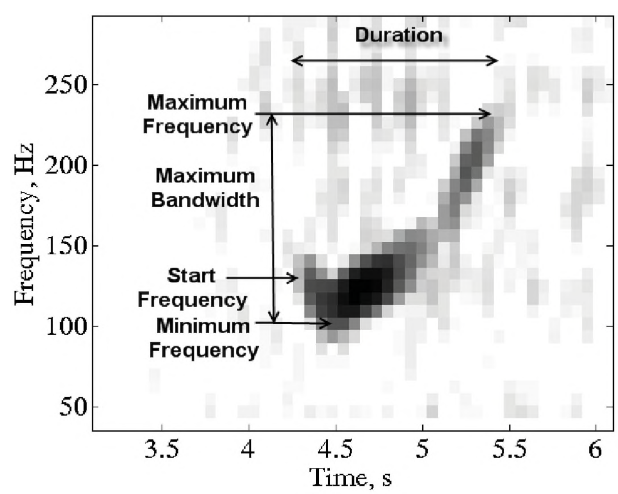

Fig. 5. Features extracted from the NARW contact call.

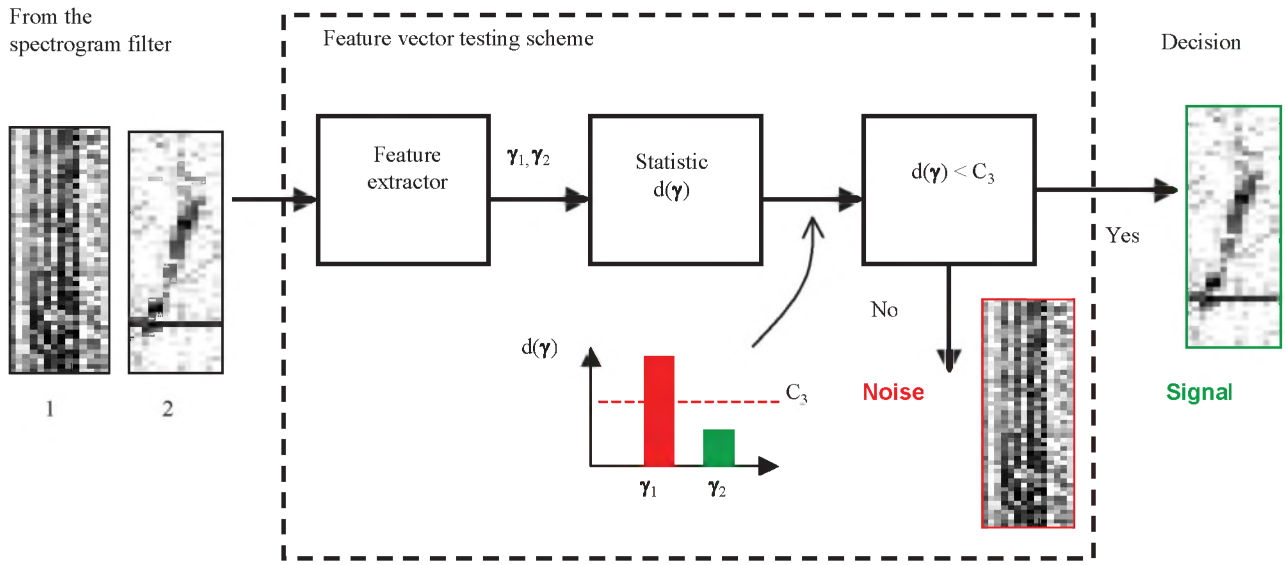

Fig. 6. Feature vector testing scheme.

calls, the FV belongs to some subspace of the K-dimensional feature space. For example, it is known that for the typical NARW contact calls, the minimum frequency is $50 \le v_4 \le 150$ Hz and the maximum bandwidth is $20 \le v_3 \le 170$ Hz. Based on the empirical observations, similar one-dimensional sets can be introduced for each element of the feature vector so that the signal subspace can be defined as:

$$V_S = \{ \boldsymbol{v} | v_{i\,min} \le v_i \le v_{i\,max},\ i = 1,2,\ldots,K \} \subset E^K \tag{2}$$

where $v_{i\,min}, v_{i\,max}$ are the scalars defining the bounds of the ith feature. As a measure of discrimination between any particular FV $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{X})$ and V_S, we introduce the following discriminant function:

$$h(\boldsymbol{v}) = \sum_{i=1}^{K} h(v_i) \tag{3}$$

where

$$h(v_i) = \begin{cases} 0 \text{ if } v_{i\,min} \le v_i \le v_{i\,max} \\ A_{i\,min}(v_i - v_{i\,min})^2 \text{ if } v_i < v_{i\,min}, \\ A_{i\,max}(v_i - v_{i\,max})^2 \text{ if } v_i > v_{i\,max}, \end{cases} \tag{4}$$

$A_{i\,min}$ and $A_{i\,max}$ are the scalars. The value $h(\boldsymbol{v})$ is compared with a threshold C_R and the hypothesis H_S is accepted if $h(\boldsymbol{v}) \le C_R$. Otherwise the hypothesis H_S is rejected. The FV testing scheme is displayed in Fig. 6. The values $v_{i\,min}, v_{i\,max}, A_{i\,min}$ and $A_{i\,max}$ specifying the FV testing algorithm are shown in Table I.

In practice, the unknown parameters $v_{i\,min}, v_{i\,max}, A_{i\,min}$ and $A_{i\,max}$ can be determined using the training data set. It is worth noting that although the proposed FV testing algorithm is heuristic, it uses the statistical properties of signals and noise. As a result, the algorithm can provide high recognition performance.

4. TEST RESULTS

Since the signal recognition technique considered here involves a two-stage decision-making process, it is difficult to estimate the recognition performance in terms of conventional receiver operating characteristics. Therefore, the performance has been evaluated using the empirical probabilities of signal recognition and false alarm.

The empirical probability of recognition has been evaluated using NARW contact calls detected by the human operators from different testing data sets. (The testing data sets were different from the training data set.) The detector and recognizer thresholds were selected as C_D = 0.35 and C_R = 1, respectively. Under these values of the threshold, the probability of signals recognition is close to 0.8. The actual probabilities of recognition obtained for different data sets are shown in Table II.

The empirical probability of false alarm has been estimated using a data set CCB04 collected in Cape Cod Bay from December 18, 2002, to January 18, 2003. Since the ambient noise conditions may change dramatically with time, the false alarm probability was computed for chunks of data 24

Table I. Parameters of the feature space

Feature	Parameters			
	$v_{i\,min}$	$v_{i\,max}$	$A_{i\,min}$	$A_{i\,max}$
Signal duration, v_1	0.5 s	1.5 s	2.5	1
Mean value of the intermediate bandwidth, v_2	0	15 Hz	0	0.03
Start-end bandwidth, v_3	20 Hz	170 Hz	0.025	0.015
Minimum frequency, v_4	65 Hz	170 Hz	0.055	0.025
Maximum bandwidth, v_5	20 Hz	170 Hz	0.05	0.015
Duration of upsweep part of the signal, v_6	0.3 s	1.5 s	4	2
Segmentation threshold, v_7	4	10	0.3	0
Local noise level, v_8	0	0.02	0	4
Percentage of holes in the object, v_9	0	0	0	3
Percentage of downsweeps in the IF, v_{10}	0	0	0	4
Percentage of harmonicas in the IF, v_{11}	0	0.3	0	3

Table II. Probability of recognition

Data set	Observation time	Number of tested signals	Probability of recognition
CCB00	08/03/2001 – 10/04/2001	14394	0.88
CCB02	28/03/2002 – 31/05/2002	1475	0.81
CCB03	21/11/2002 – 18/12/2002	67	0.79
CCB04	18/12/2002 – 18/01/2003	1331	0.83
CCB05	18/01/2003 – 04/03/2003	1792	0.85
CCB06	04/03/2003 – 21/04/2003	313	0.8
CCB09	28/02/2004 – 17/04/2004	2220	0.86
Total		21592	0.87

h in length each. The results of this test are depicted in Fig. 7. The total number of false alarms produced by the proposed technique was 826.

Test results demonstrate that the use of the proposed recognition technique essentially reduces the false alarm probability from the spectrogram-based detector [11]. In particular, the CCB04 data set contains 1,331 NARW contact calls detected by human operators (see Table II). The spectrogram-based detector was able to detect 1,289 signals so that the detection probability on the detector output was $P_D = 0.97$. The total number of false positives provided by the spectrogram-based detector was 113,341. The proposed recognizer was able to recognize 1,092 out of 1,289 signals so that the total decrease in probability of detection was $r_d = 1289/1092 = 1.18$. The corresponding decrease in probability of false alarm was $r_{fa} = 113341/826 = 137.2$.

5. DISCUSSION

For the threshold values applied, the probability of recognition of NARW contact calls ranged from 0.79 to 0.88 (see Table II). The decrease in the probability of recognition can be explained by the influence of the following factors. First, a certain number of selected calls were hardly visible on the spectrogram and hence had relatively low SNR. Although investigation of the influence of the SNR on the recognition probability was not within the scope of this work, test results demonstrate that the recognition probability decreases as the SNR goes down. Moreover, as the results reported in [5] show, when detecting the signals with low SNR, the human operators may select up to 85 data segments with no signals on them per 24 h of observed data. The operators can also make false selections because of the similarity between contact calls and some kinds of impulsive noise. In our testing data, a certain number of selections made by the operator are questionable and are not approved by other operators. The *a priori* uncertainty regarding a signal parameter is a fundamental problem in passive bioacoustics. The actual range of signal variability is unknown to the observers. As a result, there is a nonzero probability that the selection made by the human operator is actually noise. Hence, the actual probability of recognition can be higher than that represented in Table II.

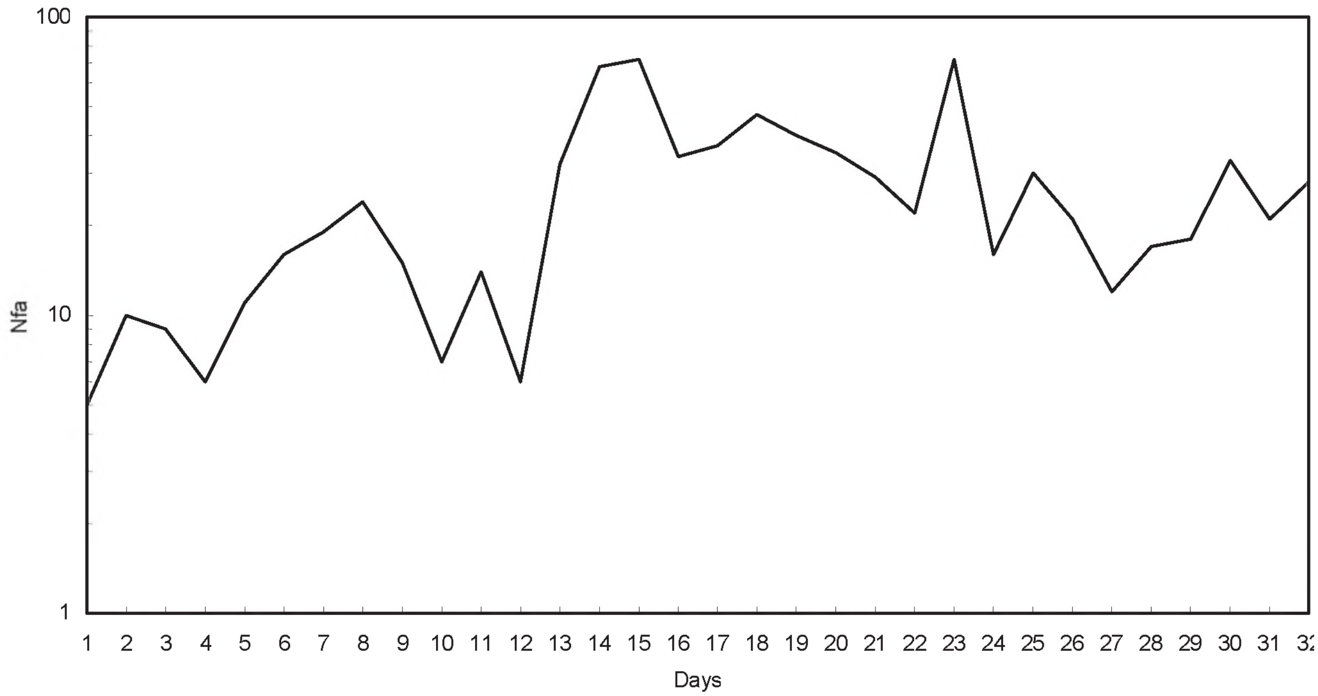

Fig. 7. The number of false alarms provided by the proposed technique. Data collected at Cape Cod Bay from December 18, 2002, to January 18, 2003.

Another factor decreasing the recognition probability is that a certain percentage of the selected NARW contact calls overlapped with transients or other calls. The proposed technique was not designed to operate under such conditions. Therefore, the problem of distinguishing partly overlapped signals should be a topic of future research.

Using the proposed technique for analyzing a long data recording, the human operator only has to inspect the automatically detected clips, thereby reducing the amount of time necessary. As test results show, an average of 26 false alarms per 24 h of observation is generated by the proposed technique. (For the duration of the false data segment equal to 1.024 s, the 26 false alarms per 24 h of observation corresponds to the false alarm probability of 3.08×10^{-4}.) Testing 26 data segments requires about 1 min per operator whereas a complete browsing of one day's worth of data by visual analysis of the spectrogram and by listening to the data requires 2–8 h per operator. Different data sets will likely have higher or lower numbers of false alarms on the recognizer output. However, practical use of the proposed technique in the Bioacoustics Research Program at the Cornell Laboratory of Ornithology shows that, on average, the human effort needed to detect more then 80% of NARW contact calls can be reduced by more than 20 times as compared with the data analysis performed by a human operator alone

6. CONCLUSION

In the presence of a high impulsive noise rate, the probability of false alarm provided by the spectrogram-based detector can increase dramatically. To decrease false alarm probability without negatively affecting the probability of detection, a new technique proposed in this paper can be used. This technique is based on a multiple-stage decision-making process involving the spectrogram and feature vector testing algorithms.

Test results demonstrate that applying the proposed signal recognition technique to the spectrogram-based detector makes it possible to reduce the false alarm probability by more than 100 times when decreasing the probability of detection by 1.2 times as compared with the spectrogram-based detector. Correspondingly, the hours that humans need to detect 80% and more of NARW contact calls can be reduced by more than 20 times as compared with the data analysis performed by a human operator alone.

7. ACKNOWLEDGEMENTS

The author wishes to thank E. Rowland and M. Fowler for their assistance in marking right whale calls and for useful discussions. Thanks also to P. Leonard for editing the draft version of the manuscript. Research funded by NOAA grant NA03NMF4720493.

8. REFERENCES

[1] Stellwagen Bank National Marine Sanctuary. Passive acoustic monitoring. Available: http://stellwagen.noaa.gov/science/passive_acoustics.html. Last viewed online 10 October 2007.

[2] Cornell Lab of Ornithology, Bioacoustics Research Program. Undersea recording: Pop-Ups. Available: http://www.birds.cornell.edu/brp/hardware/pop-ups. Last viewed online 10 October 2007.

[3] C. Clark, T. Calupca, D. Gillespie, K. Von der Heydt and J. Kemp, "A near-real-time acoustic detection and reporting system for endangered species in critical habitats," *J. Acoust. Soc. Am*. vol. 117, p. 2525, 2005.

[4] D. K. Mellinger, S. L. Nieukirk, H. Matsumoto, S. L. Heimlich, R. P. Dziak, J. Haxel and M. Fowler, "Seasonal occurrence of North Atlantic right whale (*Eubalaena Glacialis*) vocalizations at two sites on the Scotian shelf," *Marine Mammal Science*, vol. 23, pp. 856-867, Oct. 2007.

[5] I. Urazghildiiev and C. Clark, "Detection performances of experienced human operators compared to a likelihood ratio based detector," *J. Acoust. Soc. Am.*, vol. 122, pp. 200 – 204, 2007.

[6] D. K. Mellinger and C. W. Clark, 'Recognizing transient low-frequency whale sounds by spectrogram correlation," *J. Acoust. Soc. Am*. vol. 107, pp. 3518 – 3529, 2000.

[7] D. K. Mellinger, "A comparison of methods for detecting right whale calls," *Canadian Acoustics*, vol. 32, pp. 55-65, 2004.

[8] L. M. Munger, D. K. Mellinger, S. M. Wiggins, S. E. Moore and J. A. Hildebrand, "Performance of spectrogram cross-correlation in detecting right whale calls in long-term recordings from the Bering Sea," *Canadian Acoustics*, vol. 33, pp. 25-34, 2005.

[9] D. Gillespie, "Detection and classification of right whale calls using an "edge" detector operating on a smoothed spectrogram," *Canadian Acoustics*, vol. 32, pp. 39-47, 2004.

[10] I. Urazghildiiev and C. Clark, "Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test," *J. Acoust. Soc. Am.*, vol. 120, pp. 1956 – 1963, 2006.

[11] I. Urazghildiiev and C. Clark, "Acoustic detection of North Atlantic right whale contact calls using the spectrogram-based statistics," *J. Acoust. Soc. Am.*, vol. 122, pp. 769 – 776, 2007.

Photo Credit:Erin A. Falcone. Copyright: Cascadia Research, Olympia, WA, USA