Automatic grunt detector and recognizer for Atlantic cod (*Gadus morhua*)

Ildar R. Urazghildiiev[1,a)] and Sofie M. Van Parijs[2]
[1]*JASCO Applied Sciences, Suite 200, 310 K Street, Anchorage, Alaska 99501, USA*
[2]*Northeast Fisheries Science Center, Woods Hole, Massachusetts 02540, USA*

(Received 22 June 2015; revised 6 April 2016; accepted 14 April 2016; published online 9 May 2016)

Northwest Atlantic cod (*Gadus morhua*) have been heavily overfished in recent years and have not yet recovered. Passive acoustic technology offers a new approach to identify the spatial location of spawning fish, as well as their seasonal and long term persistence in an area. To date, the lack of a species-specific detector has made searching for Atlantic cod grunts in large amounts of passive acoustic data cumbersome. To address this problem, an automatic grunt detection and recognition algorithm that processes yearlong passive acoustic data recordings was designed. The proposed technique is a two-stage hypothesis testing algorithm that includes detecting and recognizing all grunt-like sounds. Test results demonstrated that the algorithm provided a detection probability of 0.93 for grunts with a signal-to-noise ratio (SNR) higher than 10 dB, and a detection probability of 0.8 for grunts with the SNR ranging from 3 to 10 dB. This detector is being used to identify cod in current and historical data from U.S. waters. Its use has significantly reduced the time required to find and validate the presence of cod grunts. © *2016 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4948569]

[WWA] Pages: 2532–2540

I. INTRODUCTION

North Atlantic cod (*Gadus morhua*) is a demersal predatory fish that has been targeted by both commercial and recreational fisheries for centuries (Lear, 1998). In the northwest Atlantic, cod was heavily overfished throughout its range, which resulted in a population crash in several U.S. and Canadian stocks during the early 1990s (Serchuk and Wigley, 1992; Fogarty and Murawski 1998; Frank *et al.*, 2011). Stock declines are frequently accompanied by reduced spawning diversity and population structural collapse of the species. In the Canadian Scotian Shelf ecosystem, both predator and prey species appear to be reverting to historic levels (Frank *et al.*, 2011). However, in the U.S., concerns about stock recovery remain. Zemeckis *et al.* (2014) stressed that research that focuses on Atlantic cod spawning can provide insights into the mechanisms involved in rebuilding stock. The most important considerations for fisheries managers are the spatial and temporal extent of cod spawning, the behavior of spawners, and the reproductive contributions of older age classes.

Passive acoustic monitoring is a well-recognized tool for collecting a wide range of biological information about marine animals, including fish species (e.g., Zelick *et al.*, 1999; Moore *et al.*, 2006; Rountree *et al.*, 2006; Mellinger *et al.*, 2007; Luczkovich *et al.*, 2008; Slabbekoorn *et al.*, 2010; Wall *et al.*, 2013). Because passive acoustic monitoring uses acoustic recorders that routinely collect data for months or years (e.g., Rowell *et al.*, 2012; Schärer *et al.*, 2012; Hernandez *et al.*, 2013), it is unaffected by time of day or most weather conditions, and is thus highly versatile. This approach has been shown to provide detailed information on the spatial and temporal extent of spawning aggregations of fish (e.g., Hernandez *et al.*, 2013; Wall *et al.*, 2013; Wall, 2014).

Atlantic cod produce low-frequency grunts—broadband signals that range from 50 to 100 Hz (e.g., Hawkins and Rasmussen, 1978; Finstad and Nordeide, 2004). Compared to other gadoid fishes, cod have a limited vocal repertoire; they produce grunts in multiple social contexts (Finstad and Nordeide, 2004; Rowe and Hutchings, 2006). Early studies conducted on fish in tanks clearly demonstrated that males grunted during the spawning season as part of their courtship display (Brawn, 1961a,b,c). Studies on captive and wild cod found that cod grunting peaked with spawning activity (Nordeide and Kjellsby, 1999; Rowe and Hutchings, 2006; Hernandez *et al.*, 2013).

Cod produce a range of sounds other than grunts: "knocks" produced singularly when they interact with members of their own species (Midling *et al.*, 2002), a high frequency (~2 kHz) "click" produced when potential predators are present (Vester *et al.*, 2004), and a low-frequency "hum" made during the ventral mount (Rowe and Hutchings, 2006). Because the knocks, hums, and clicks were identified in recordings of cod in captivity, the algorithm development was focused on detecting the Atlantic cod "grunt" that has been clearly linked to spawning activity in the wild.

Because passive acoustic technology used to study spawning aggregations generates large quantities of acoustic data over a short time, data processing takes a long time, which hampers expedient management decisions. For Atlantic cod, the lack of an automated detector has significantly

[a)]Electronic mail: ildar.urazghidiiev@jasco.com

 © 2016 Acoustical Society of America

impeded processing current and historical passive acoustic data, which could help identify spawning aggregations. Hernandez *et al.* (2013) investigated passive acoustic data records that were highly subsampled to extract relevant information within a reasonable time. Although subsampling—a process in which only part of the sound recordings are analyzed—is a justifiable approach, it suffers from being able to only examine a small portion of information, thus, potentially misses important events. Additionally, subsampling is not an effective approach to search for new spawning aggregations in thousands of hours of passive acoustic data. Therefore, developing an Atlantic cod grunt detector was clearly a critical next step.

This paper addresses the problem of passive acoustic detection of grunts produced by Atlantic cod. The main goal was to design an automatic grunt detection and recognition algorithm that could process yearlong passive acoustic data records quickly enough to reduce the hours a person would otherwise spend analyzing large volumes of data. The algorithm was divided into a two-stage hypothesis testing technique. In the first stage, the spectrogram-based detector scanned data recordings, during which time it tagged all grunts and grunt-like sounds. In the second stage, a feature vector testing algorithm recognized the detected events and evaluated the validity of each tagged event.

The basic requirement for the automatic grunt detection and recognition algorithm designed for use in this work was that it detect sounds similar to how a human analyst would. This requirement was made because in most applications an analyst makes the final decision whether grunts were present or absent and rejects signals that were so weak as to not show on spectrograms (Urazghildiiev and Clark, 2007b). Therefore, we deemed it important that the automatic algorithm should only detect signals that show on a spectrogram. The detector's accuracy was evaluated against data in which experienced acoustic analysts had marked definite grunts.

II. DATA MODEL AND PROBLEM FORMULATION

Atlantic cod produce low-frequency impulsive signals called "grunts" (Fig. 1).

As Fig. 1 shows, most of each signal's energy is distributed over a small number of harmonics, such that the signal power spectrum density (PSD) can be modeled as a sum of M components

$$S(f,t) = \sum_{m=1}^{M} S_m(f,t), \quad t \in [0,T], \quad f \in [F_{\min}, F_{\max}], \tag{1}$$

where $S_m(f,t)$ is the function specifying the PSD of the mth harmonic, T is the duration of the signal, and $F_{\min}, F_{\max}$ are the bounds of the frequency range of grunts. For all harmonics, the assumption is $F_{\min} = 20$ Hz and $F_{\max} = 500$ Hz. Figure 1 also shows that each harmonic can be represented as a locally narrowband frequency-modulated signal. To model the PSD of each mth harmonic, the exponential kernel is proposed

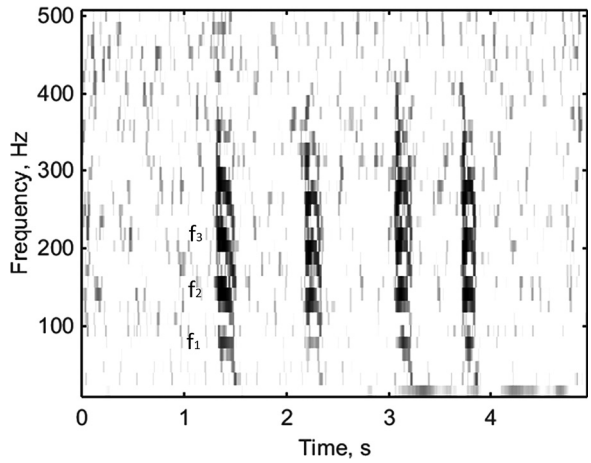

FIG. 1. Spectrograms of four grunts produced by Atlantic cod. The spectrogram was created using 128 point fast Fourier transform (FFT) of acoustic recordings with sampling rate 2 kHz, frequency resolution 7.8125 Hz, Hanning window, and 50% overlap.

$$S_m(f,t) = p_m \exp\left\{-\frac{(f - f_m(t))^2}{2B^2}\right\}, \tag{2}$$

$$f_m(t) = mf_1 + \alpha_m t, \tag{3}$$

where p_m is the coefficient proportional to the energy of the mth harmonic, B is the bandwidth, f_1 is the start frequency of the first harmonic, $\alpha_m = \Delta_{fm}/T$ is the frequency slope of the mth harmonic, and Δ_{fm} is the frequency deviation. Because the analysis showed that differences in frequency deviations and slopes of harmonics do not significantly affect how algorithms perform, we assumed the bandwidth and frequency deviations were the same for all harmonics, such that $\alpha_m = \alpha$ for all $m = 1, \ldots, M$. Figure 1 shows that the bulk of signal energy is distributed over three harmonics with the frequencies $f_1 < f_2 < f_3$. Therefore, our model was restricted by using $M = 3$ harmonics in Eq. (1).

Stemming from the proposed models (1)–(3), distribution of signal energy in the time-frequency plane can be specified by the random vector of parameters $\boldsymbol{\lambda} = [f_1, f_2, f_3, \Delta_f, T, B]^T$, such that $S(f,t) = S(\boldsymbol{\lambda})$.

The acoustical sensor provided digitized series of real-valued samples with a sampling rate greater than twice the highest frequency in Atlantic cod grunts, $F_{\max}$. For each subset of data starting at time t, the whitened spectrogram, $X(f,t) \in E^{N_f \times N_t}$, $f \in [F_{\min}, F_{\max}]$, normalized in the units of PSD, can be computed (see Urazghildiiev and Clark, 2007a, for details). Here, N_f and N_t are the number of frequency bins and time slices of the spectrogram, respectively. The following hypotheses were introduced

$$H_0 : X(f,t) = W(f,t);$$

$$H_S : X(f,t) = S(\boldsymbol{\lambda}) + W(f,t);$$

$$H_I : X(f,t) = Q(f,t) + W(f,t);$$

where $W(f,t)$ is the spectrogram of continuous ambient noise, and $Q(f,t)$ is the spectrogram of impulsive noise. Impulsive noise is any process that occupies the same

frequency band $[F_{\min}, F_{\max}]$ as grunts, has a similar duration, and satisfies the condition $|Q(f,t)| < \infty$, where the symbol $|\cdot|$ denotes matrix norm. Although *a priori* probabilities of the hypotheses, $p(H_0)$, $p(H_S)$, and $p(H_I)$ are not known, grunts and impulsive noise are relatively rare events; therefore, it was assumed that the following conditions hold true:

$$p(H_0) \gg p(H_S), \quad p(H_0) \gg p(H_I). \tag{4}$$

The final assumption is that a human operator, hereafter referred to as an analyst, would verify all signals automatically detected and score them as correct or incorrect for a final quality check. The problem can be formulated to accept or reject the hypothesis H_S that the spectrogram $X(f,t)$ contains an Atlantic cod grunt.

III. HYPOTHESIS TESTING ALGORITHMS

The analyst decided whether grunts were present or absent, concurrently rejecting signals that were so weak they did not show on the spectrograms (Urazghildiiev and Clark, 2007b). To make the automatic grunt detection and recognition algorithm perform similar to an analyst, it should only detect signals that are visible on the spectrogram; thus, the class of acceptable detectors was restricted to those that use spectrograms to compute detection statistics.

The general strategy to solve the grunt detection problem was determined by considering the assumption (4). Because the hypothesis H_0 has the highest *a priori* probability, the automatic grunt detection and recognition algorithm based on a two-stage hypothesis testing technique was proposed. In the first stage, called signal detection, the hypothesis H_S was tested against the null hypothesis H_0. If, for a segment of tested data, the hypothesis H_S was rejected, no action was taken and the next portion of data was tested. However, if the hypothesis H_S was not rejected, the second stage of testing the hypothesis H_S versus H_I began; this stage is called signal recognition.

A. Signal detection

Based on the models (1)–(3), the decision-making algorithm similar to that used to detect North Atlantic right whale contact calls (Urazghildiiev *et al.*, 2009) is proposed. A bank of two-dimensional (2-D) linear filters for computing the detection statistic

$$z(t) = \max_k |u(t, \boldsymbol{\lambda}_k)|^2 \tag{5}$$

was used, where

$$u(t, \boldsymbol{\lambda}_k) = X(f,t) * S(\boldsymbol{\lambda}_k) \tag{6}$$

is the output of the kth linear filter, and $S(\boldsymbol{\lambda}_k)$ is the kernel of the kth linear filter computed according to Eqs. (2) and (3) for a particular vector $\boldsymbol{\lambda}_k$, and the asterisk, "*," denotes bilateral convolution. The statistic $z(t)$ was compared with a threshold, C, and the hypothesis H_S was accepted for $X(f,t)$ if $z(t) \geq C$. The analyst applies the threshold based on the requirement to detect signals on the spectrogram. In this work, an adaptive threshold $C = 4z_{\text{med}}$ was used, where z_{med} is the median value of $z(t)$ computed over 30 s. Figure 2 shows the filter output, $z(t)$, computed using a $k = 1$ linear filter for the chunk of data displayed in Fig. 1. The time instances, $\widehat{t}_i$, $i = 1, \ldots, 4$, corresponding to the peak values of $z(t) \geq C$, are taken as time of arrival (TOA) estimates of the detected signals. The next stage tests the spectrograms when H_S is accepted.

B. Signal recognition

The proposed signal recognition algorithm is based on tests of the features extracted from the spectrogram $X(f,t)$.

1. Feature extraction

As Fig. 1 shows, the most prominent visual feature of Atlantic cod grunts is the presence of three or more harmonics separated by 50–80 Hz in the frequency domain. Each harmonic has an approximately equal slope, α. Figure 3(a) shows a spectrogram of a grunt whose harmonics start at $f_1 = 156.3$ Hz, $f_2 = 226.6$ Hz, and $f_3 = 289.1$ Hz. To extract these features, two transformations of the spectrogram are proposed.

The first transformation is

$$S_1(t) = \sum_{f=1}^{N_f} \tilde{X}(f,t), \tag{7}$$

where $\tilde{X}(f,t) = 10 \log_{10} X(f,t)$ is the spectrogram represented in dB. Plot of the function $S_1(t)$ is shown in Fig. 3(c). Using $S_1(t)$, signal duration was estimated by

$$\widehat{T} = t_{\max} - t_{\min}, \tag{8}$$

where $t_{\min}$ and $t_{\max}$ are the start and end times of the spectrogram in which the function $S_1(t)$ exceeds $0.3S_{1\max}$, and $S_{1\max}$ is the peak of $S_1(t)$ taken near the TOA estimate, $\widehat{t}$, of the detected signal. Note that in Eq. (7), a normalized spectrogram was used. For each $f \in [F_{\min}, F_{\max}]$, the normalized spectrogram of continuous background noise has the property

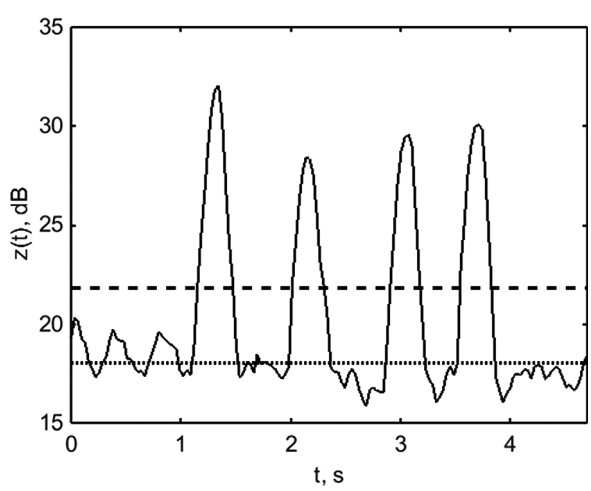

FIG. 2. The output of a linear filter (5), $z(t)$, the threshold, C (dashed line), and the value of z_{med} (dotted line).

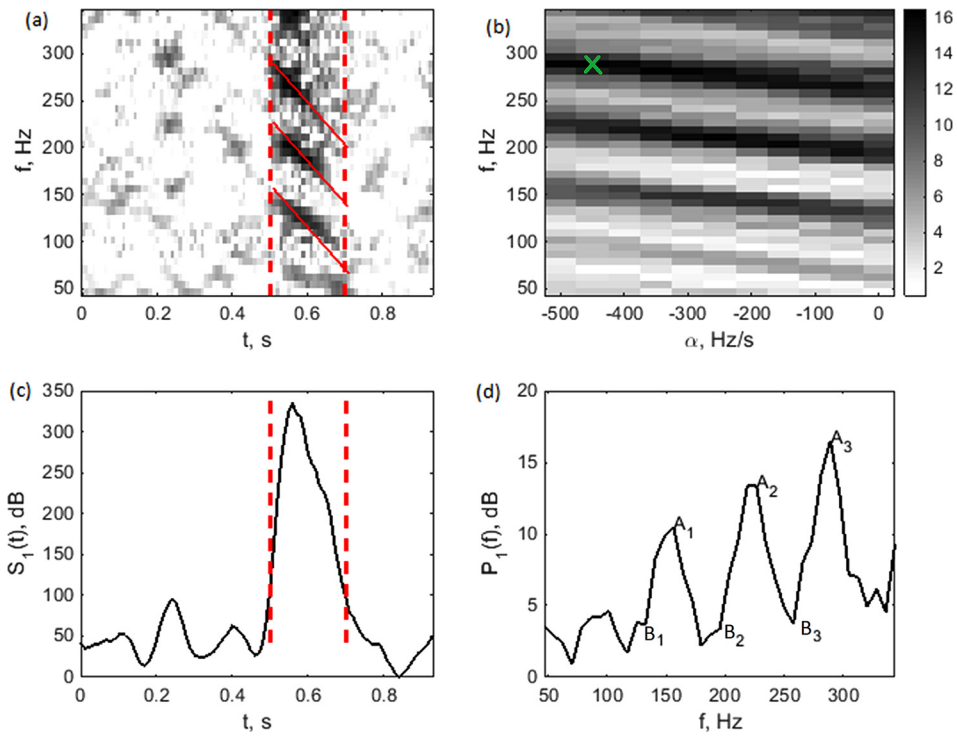

FIG. 3. (Color online) (a) Spectrogram of grunt. Dotted lines estimate start and end time; solid lines estimate slope. (b) Function $P(f,\alpha)$ computed for spectrogram of grunt. The gray scale represents the value of $P(f,\alpha)$, dB, falling within each frequency-slope bin. The symbol "×" indicates the peak frequency and the slope. (c) Function $S_1(t)$. (d) Function $P_1(f)$.

$$\text{median}\{W(f,t=1,\ldots,N_t)\}=1,$$

such that

$$\sum_{f=1}^{N_f}\sum_{t=1}^{N_t}W(f,t)=N_fN_t,$$

where N_f is the number of frequency slices of the spectrogram within the frequency interval $f\in[F_{\min},F_{\max}]$ and $N_t\gg 1$ is the number of time slices used to compute a normalized spectrogram. Thus, the signal-to-noise ratio (SNR) estimate is computed by

$$\widehat{\text{SNR}}=\frac{1}{N_fN_{ts}}\sum_{f=1}^{N_f}\sum_{t=t_{\min}}^{t_{\max}}X(f,t), \tag{9}$$

where N_{ts} is the number of time slices of the spectrogram within the time interval $t\in[t_{\min},\,t_{\max}]$.

The second transformation is based on the 2-D function

$$P(f,\alpha)=\sum_{\tau=\hat{t}-0.05}^{\hat{t}+0.05}\tilde{X}(f+\tau\alpha,\tau) \tag{10}$$

computed for the time interval $\hat{t}\pm 0.05$ s. For $\alpha=0$, the function $P(f,0)$ computes the mean value of the fast Fourier transform (FFT) spectrums generated from the data and represents the PSD estimate. For $\alpha\neq 0$, the function $P(f,\alpha)$ computes the mean values of the FFT spectrums over the frequency slope α.

Using the transformation (10), the estimates of the peak frequency of the signal and its frequency slope can be computed as

$$\{\hat{f}_P,\hat{\alpha}\}=\arg\max_{f,\alpha}P(f,\alpha). \tag{11}$$

Figure 3(b) is a plot of the function $P(f,\alpha)$ computed for the frequency grid $f\in[46.9,\,46.9+\Delta_f,\ldots,343.8]$ Hz, $\Delta_f=7.8$ Hz, and for the slope grid $\alpha\in[-500,-450,\ldots,0]$ Hz/s. In Fig. 3(b), "×" denotes the estimates $\{\hat{f}_P,\hat{\alpha}\}$.

The one-dimensional (1-D) generalized PSD estimate is computed from Eq. (10) as

$$P_1(f)=P(f,\hat{\alpha}). \tag{12}$$

Figure 3(d) shows a plot of the function $P_1(f)$. This plot shows three highest peaks, A_1, A_2, and A_3, located at the start frequencies, $f_1<f_2<f_3$, of the corresponding signal harmonics. The averaged inter-harmonic interval (IHI) was estimated as

$$\hat{\delta}_f=0.5((f_2-f_1)+(f_3-f_2)).$$

An important property of grunts is that there is relatively little energy between harmonics. This low energy corresponds to low values of $P_1(f)$ between the highest peaks. If B_m, $m=1,\ldots,3$, is the maximum value of $P_1(f)$ at frequencies $0.2\hat{\delta}_f\le|f-f_m|\le 0.5\hat{\delta}_f$ around the mth peak, then the relative amount of energy around the mth harmonic can be specified by the ratio

$$r_m=\frac{A_m}{B_m}.$$

The following features, which were extracted from $S_1(t)$ and $P_1(f)$, were used to recognize signals

$$x_1=\hat{T}(\text{duration}),$$
$$x_2=\hat{f}_P(\text{peak frequency}),$$
$$x_3=\widehat{\text{SNR}}(\text{SNR}),$$
$$x_4=\hat{\delta}_f(\text{IHI}),$$
$$x_5=\min_{m=1.3}r_m(\text{peak-to-min ratio}),$$
$$x_6=\frac{A_{\max}}{A_{\min}}(\text{peak-to-peak ratio}),$$

where $A_{\max}$ and $A_{\min}$ are the maximum and minimum values of the peaks $\{A_1, A_2, A_3\}$, respectively.

2. Feature testing

The random feature vector, $\boldsymbol{x}(\hat{t}) = \boldsymbol{x} = [x_1, \ldots, x_6]^T$, extracted from the detected signal, was the deciding factor in accepting or rejecting the hypothesis H_S. The design of the decision-making algorithm was based on a well-developed maximum likelihood approach that provides acceptable practical solutions in many applications.

The maximum likelihood algorithm requires knowing the probability density functions (PDFs), $W(\boldsymbol{x}|H_S)$, $W(\boldsymbol{x}|H_I)$, of $\boldsymbol{x}$ under the hypothesis H_S and the alternative hypothesis H_I. Because no assumptions about $W(\boldsymbol{x}|H_I)$ were made, the decision-making algorithm was reduced to comparing $W(\boldsymbol{x}|H_S)$ with a threshold. If

$$W(\boldsymbol{x}|H_S) \geq C_x, \tag{13}$$

then the hypothesis H_S is accepted. The features $x_1, \ldots, x_6$ were assumed to be independent random variables. In this case, the PDF is factorized as

$$W(\boldsymbol{x}) = W(x_1)W(x_2)\cdots W(x_6). \tag{14}$$

The dependence of PDF on the hypothesis H_S was omitted. In many applications, $W(x_n)$ are assumed to be Gaussian. PDFs of the Gaussian-distributed feature are represented as

$$W(x_n) = k_n \exp\{-d_n^2\}, \tag{15}$$

$$d_n^2 = \frac{(x_n - m_n)^2}{2\sigma_n^2}, \tag{16}$$

where m_n is the mean value of x_n, σ_n^2 is its variance, and k_n is a scalar. Because k_n does not depend on x_n, this factor was not considered. Substituting Eqs. (15) and (16) into Eq. (14) gives

$$W(\boldsymbol{x}) = \exp\left\{-\sum_{n=1}^{6} d_n^2\right\}, \quad W(\boldsymbol{x}) \in [0,\ 1]. \tag{17}$$

The global maximum of Eq. (17) is achieved when $x_n = m_n$ for all $n = 1, \ldots, 6$,

$$W_{\max} = W(x_n = m_n,\ n = 1, \ldots, 6) = 1.$$

In (17), d_n are weighting functions. Following from Eq. (16), the lowest weight that equals zero is assigned to $x_n = m_n$. In general, the closer x_n is to its mean value, m_n, the lower d_n and the higher PDF $W(\boldsymbol{x})$ is. Assigning low weights to the most probable values can be generalized by assigning low weights to the most acceptable values. For example, because grunt durations can vary from 57 to 360 ms (Hernandez *et al.*, 2013), the durations from 120 to 250 ms are equally acceptable. Hence, an equally low weight can be assigned to any $x_1 \in [120,\ 250]$ ms.

The approach based on subjective probabilities was used in this work to design the grunt recognition algorithm. Subjective PDF of the nth feature was defined in the form (15) using the following weighting function:

$$d_n^2 = \begin{cases} 0,\ x_n \in [m_{n1}, m_{n2}] \\ (x_n - m_{n1})^2/2\sigma_{n1}^2,\ x_n < m_{n1} \\ (x_n - m_{n2})^2/2\sigma_{n2}^2,\ x_n > m_{n2}, \end{cases} \tag{18}$$

where m_{n1}, m_{n2} are the lower and upper bounds, respectively, of the interval that corresponds to zero weight and σ_{n1}^2, σ_{n2}^2 are the parameters that specify the increase of the weight outside of the interval $[m_{n1}, m_{n2}]$. PDF (15) with the weighting function (18) is a generalized Gaussian distribution (16). The advantage of the proposed approach is its flexibility. The parameters that specify the weighting function (18) can be defined subjectively based on the developer's knowledge about the selected feature, empirical distributions of corresponding features, and practical needs.

Thus, the proposed signal recognition step was condensed to extracting the feature vector, $\boldsymbol{x}$, using the functions $S_1(t)$ [Eq. (7)], $P(f, \alpha)$ [Eq. (10)], $P_1(f)$ [Eq. (12)], and comparing the subjective PDF [Eq. (17)] with a threshold (13), where the subjective PDF is defined by weighting functions (18).

IV. TEST RESULTS

A. Acoustic recordings

Algorithm performance was evaluated with data collected from a passive acoustic monitoring system that consisted of 19 marine autonomous recording units (MARUs) deployed in the eastern part of Massachusetts Bay near Stellwagen Bank (west of 70°00′W and between 42°05′N and 42°40′N north of Cape Cod Bay). The MARUs were ~9.3 km apart and placed on the sea floor in a hexagonal grid array (see Morano *et al.*, 2012, for more details). Each MARU had a flat frequency response between 15 and 585 Hz with sensitivity of −151 dB re 1 V/μPa. All MARUs recorded continuously at a sampling rate of 2 kHz. The tests were based on acoustic recordings made on November 25, 2009, October 26, 2010, November 22, 2010, and December 1, 2010.

B. Empirical distributions of signal parameters

A training data set was manually created from 3047 grunts that had SNRs ≥3 dB that were clearly visible on the spectrogram made from the December 1, 2010 data set. The empirical distributions of signal duration, T, peak frequency, f_P, IHI, δ_f, and frequency slope, α, are shown in Fig. 4.

C. Detector and recognizer parameters

The detector (5) with a single linear filter was chosen based on empirical distributions of signal parameters (Fig. 4). The parameters (6) that specify the linear filter were selected as $\boldsymbol{\lambda} = [f_1, f_2, f_3, \Delta_f, T, B]^T$, were $f_1 = 85$ Hz, $f_2 = 156$ Hz, $f_3 = 225$ Hz, $\Delta_f = -150$ Hz, $T = 0.2$ s, $\alpha = -300$ Hz/s, and $B = 15$ Hz.

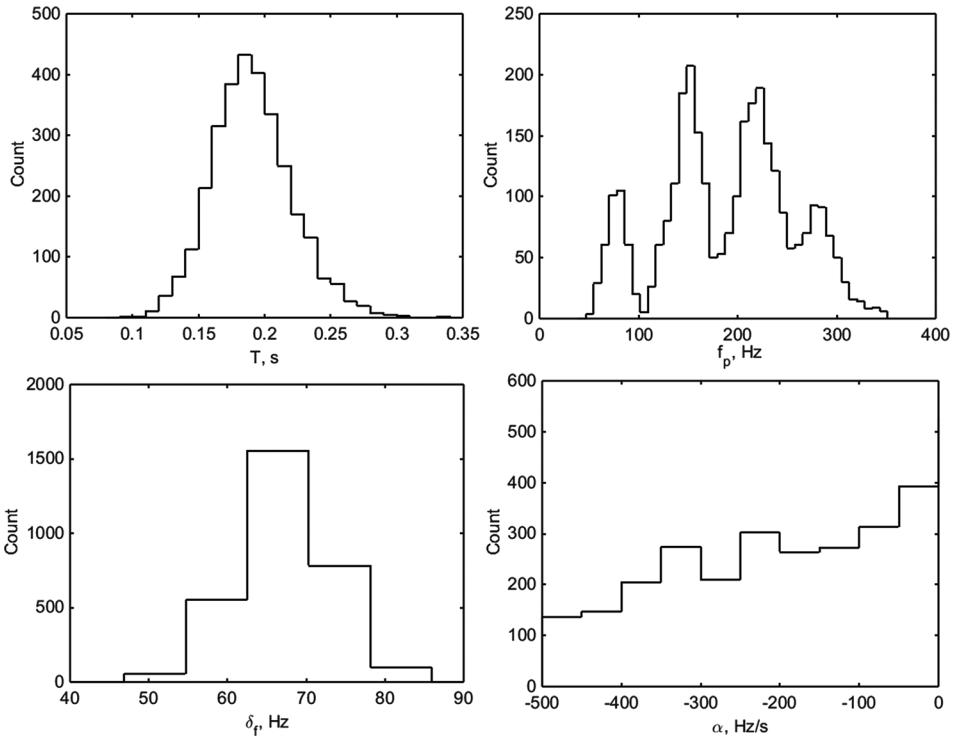

FIG. 4. (a) Empirical distributions of signal duration, T; (b) Peak frequency, f_P; (c) IHI, δ_f; (d) Frequency slope, α.

Table I lists recognizer parameters that specify weighting functions (18). Plots of PDFs of some features are shown in Fig. 5. $C_x = e^{-1} = 0.37$ was the recognizer threshold used in Eq. (13). The recognizer parameters and the threshold were chosen to minimize the probability of false detections and to provide a recognition performance similar to what a human analyst would provide.

D. Detector and recognizer performance

The test data set consisted of 24 h of acoustic recordings made on November 25, 2009. Detector (5) and recognizer (13) performances were evaluated in terms of the probabilities of signal detection, P_d, false alarm, P_{fa}, and the area under the receiver operating characteristic curve (AUC). Table II shows the results of this evaluation. The column N_{tot} represents the total number of signals and impulsive noise events manually and automatically detected in the first stage. The column N_{sig} lists the number of grunts detected manually, the column N_d lists the number of grunts correctly detected and recognized automatically, and the column P_d lists the detection probability, $P_d = N_d/N_{\text{sig}}$. The column N_{fa} represents the number of false alarms produced by the automatic algorithm; the column P_{fa} lists false alarm probability, $P_{\text{fa}} = N_{\text{fa}}/N_{\text{noise}}$.

TABLE I. Recognizer parameters that specify weighting functions (18) (N/A, not applicable).

Feature	m_{n1}	m_{n2}	σ_{n1}	σ_{n2}
x_1 (duration)	0.12 s	0.25 s	0.032 s	0.042 s
x_2 (peak frequency)	55 Hz	∞	5.7 Hz	N/A
x_3 (SNR)	5 dB	∞	2.8 dB	N/A
x_4 (IHI)	55 Hz	78 Hz	7.1 Hz	5.6 Hz
x_5 (peak-to-min ratio)	6 dB	∞	2.1 dB	N/A
x_6 (peak-to-peak ratio)	0 dB	8 dB	1 dB	2.8 dB

V. DISCUSSION AND CONCLUSIONS

Since the actual number of grunts and their parameters were not known, it was not possible to objectively evaluate detection performance with field recordings. Instead, the analyst examined the spectrogram to evaluate the number N_{sig}, and TOAs of grunts used in the test data set, a process that introduced a degree of subjectivity to the performance evaluation. Because the analyst's ability to detect signals decreases alongside the SNR, the analyst could only accurately detect signals that had sufficiently high SNRs. The detector parameters were selected so the detector's performance would more closely approximate the analyst's ability.

The most important characteristic of the proposed cod detector is its ability to detect signals with high SNRs. As Table II shows, the detection probability was 0.93 for signals with SNR $>$ 10 dB and 0.8 for signals with SNR $\in [3, 10]$ dB.

When P_d decreased, it was attributed to one of these main reasons

- Low SNR.
- When two or more signals were separated on the time axis by $<$0.1 s, they were considered to be one signal.
- When signals with impulsive noise overlapped, there was a decrease in the accuracy by which feature vectors could be estimated.
- Grunts with atypical characteristics were detected. This included grunts with low intensity of one of the harmonics decreasing peak-to-peak ratio, $W(x_6)$, grunts with low duration decreasing $W(x_1)$, and grunts with increasing frequency slope $\alpha > 0$.

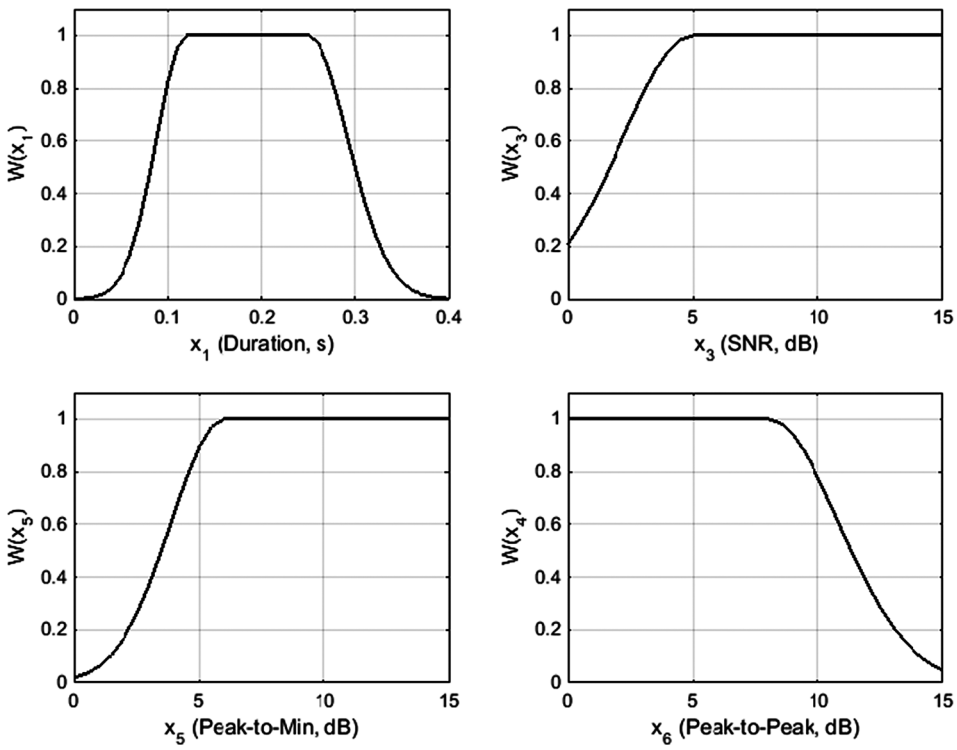

FIG. 5. Subjective probability density function (PDF) of the features: (a) Duration, x_1. (b) SNR, x_3. (c) Peak-to-min ratio, x_5. (d) Peak-to-peak ratio, x_6.

To reduce the negative effects of SNR on detection probability we specified the subjective weighting function for SNR so the PDF value would decrease when SNR < 5 dB and so automatic detection would not occur when SNR ≤ 1 dB. As Table II shows, for signals with SNR < 3 dB, the automatic detector provided the detection probability $P_d = 0.42$.

The number of false detections a detector generates is its second most important characteristic because an analyst must invest a lot of time to remove them. The detector threshold was selected to minimize the probability of false detections. There were only 10–236 false detections per 24 h of recordings (Table II). When the number of false detections is so small, it takes little time for an analyst to hand browse the data, which is the advantage of the proposed algorithm.

TABLE II. Performance of the automatic detection and recognition algorithm (N/A, not applicable).

Data set	N_{tot}	N_{sig}	N_{noise}	N_d	P_d	N_{fa}	P_{fa}	AUC
MARU 1	9722	7353	2369	5625	0.76	26	0.011	0.9614
MARU 2	940	8	932	8	1	17	0.0182	0.9983
MARU 3	1175	18	1157	18	1	30	0.0259	0.9991
MARU 4	2096	0	2096	0	N/A	126	0.0601	N/A
MARU 5	1390	0	1390	0	N/A	27	0.0194	N/A
MARU 6	1744	22	1722	22	1	41	0.0238	0.9972
MARU 8	2110	6	2104	6	1	60	0.0285	0.9998
MARU 9	398	0	398	0	N/A	40	0.1005	N/A
MARU 10	3105	0	3105	0	N/A	166	0.0535	N/A
MARU 12	561	0	561	0	N/A	57	0.1016	N/A
MARU 13	1766	0	1766	0	N/A	197	0.1116	N/A
MARU 15	3554	0	3554	0	N/A	236	0.0664	N/A
MARU 16	1463	15	1448	15	1	29	0.02	0.9988
MARU 17	3116	3	3113	3	1	224	0.072	0.9752
MARU 19	298	0	298	0	N/A	10	0.0336	N/A
SNR < 3 dB	15 057	905	14 152	382	0.42	118	0.0083	0.9478
SNR 3,...,10 dB	15 751	5710	10 041	4564	0.8	874	0.087	0.9356
SNR > 10 dB	2630	810	1820	751	0.93	294	0.1615	0.9559
Total	33 438	7425	26 013	5697	0.77	1286	0.0494	0.9511

The number of grunts manually detected on different MARUs of the test data set varied from 0 to 7353 (Table II). The MARU 1 data set indicated a large aggregation of Atlantic cod near this recorder. Because many cod were far from MARU 1, many grunts had SNRs that were too low to detect. Figure 6 shows the empirical distributions of the peak frequencies computed as percentages of the total number of the detected signals and noise. For MARU 1 data set, when a large aggregation of cod was detected, the empirical distributions of peak frequencies of the detections manually labeled as "grunt" and "noise" have similar shapes. These distributions, however, have different shapes for MARU 8 data when only six grunts were detected. A closer examination of MARU 1 data explained this divergent distribution: A large portion of grunts had been incorrectly labeled as "noise." Because the distance between the cod and the MARU 1 was large and the SNR was low, the analyst was unable to differentiate these detections on the spectrogram. Figure 6 demonstrates that automatic detector sensitivity was higher than the analyst's ability to detect grunts from the spectrogram, a phenomenon similar to results for low-frequency sounds produced by North Atlantic right whales (Urazghildiiev and Clark, 2007b).

Empirical distributions of the grunt durations and peak frequencies (see Fig. 4) are similar to other studies (Finstad and Nordeide, 2004; Fudge and Rose, 2009; Hernandez *et al.*, 2013).

By analyzing the distribution of energy in the time-frequency plane, a grunt detection and recognition algorithm can be developed. The most prominent feature of grunts is

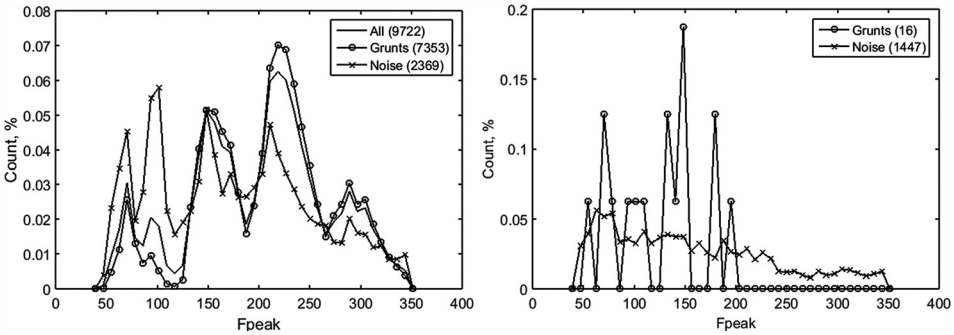

FIG. 6. Empirical distributions of the peak frequencies computed for MARU 1 (left) and MARU 8 (right) on November 25, 2009 data set.

that they all have an approximately equal slope, α, and three or more harmonics separated by 50–70 Hz in the frequency domain.

To detect grunts automatically, a two-stage hypothesis testing technique was developed. In the first stage, referred to as signal detection, the hypothesis H_S was tested versus the hypothesis H_0. In the second stage of signal recognition, the hypothesis H_S was tested against the alternative hypothesis H_I.

The grunt recognition algorithm used six measurable features, which were extracted using the functions $S_1(t)$ [Eq. (7)] and $P_1(f)$ [Eq. (12)]. The decision-making algorithm was designed using a statistical approach based on subjective probabilities. Although the PDFs were specified subjectively, they were based on empirical distributions of signal parameters. Flexibility is an important advantage of using subjective probabilities when designing an automatic recognition algorithm. Because it is fairly simple to change the subjective PDFs, new information about signal parameters is easy to incorporate, thus, readily addressing practical needs.

The proposed automatic grunt detection algorithm provides high detection probability for signals with high SNR; it dramatically decreases the hours an analyst needs to invest to thoroughly analyze data.

ACKNOWLEDGMENTS

The authors thank Annamaria Izzi and Samara Haver for helping annotate the training data set and providing helpful input, and Katherine Williams for helping edit the draft version of the manuscript. This work would not have been possible without the funding provided by NOAA Saltonstall-Kennedy program and the collaborative support of Micah Dean and Bill Hoffman at the Massachusetts division of Marine Fisheries, Doug Zemeckis at the School for Marine Science and Technology (*SMAST*) at the University of Massachusetts Dartmouth, Chris McGuire at the Nature Conservancy, Leila Hatch from Stellwagen Bank National Marine Sanctuary, and Chris Clark and Aaron Rice at Cornell University.

Brawn, V. M. (**1961a**). "Aggressive behaviour in the cod (*Gadus callarias L.*)," Behaviour **18**, 107–147.

Brawn, V. M. (**1961b**). "Reproductive behaviour of the cod (*Gadus callarias L.*)," Behaviour **18**, 177–198.

Brawn, V. M. (**1961c**). "Sound production by the cod (*Gadus callarias L.*)," Behaviour **18**, 239–255.

Finstad, J. L., and Nordeide, J. T. (**2004**). "Acoustic repertoire of spawning cod, *Gadus morhua*," Environ. Biol. Fishes **70**, 427–433.

Fogarty, M. J., and Murawski, S. A. (**1998**). "Large-scale disturbance and the structure of marine systems: Fishery impacts on Georges Bank," Ecol. Appl. **8**, S6–S22.

Frank, K. T., Petrie, B., Fisher, J. A. D., and Leggett, W. C. (**2011**). "Transient dynamics of an altered large marine ecosystem," Nature **477**, 86–91.

Fudge, S. B., and Rose, G. A. (**2009**). "Passive- and active-acoustic properties of a spawning Atlantic cod (*Gadus morhua*) aggregation," ICES J. Mar. Sci. **66**, 1259–1263.

Hawkins, A. D., and Rasmussen, K. J. (**1978**). "The calls of gadoid fish," J. Mar. Biol. Assoc. U. K. **58**, 891–911.

Hernandez, K. M., Risch, D., Cholewiak, D. M., Dean, M. J., Hatch, L. T., Hoffman, W. S., Rice, A. N., Zemeckis, D., and Van Parijs, S. M. (**2013**). "Acoustic monitoring of Atlantic cod (*Gadus morhua*) in Massachusetts Bay: Implications for management and conservation," ICES J. Mar. Sci. **70**, 628–635.

Lear, W. H. (**1998**). "History of fisheries in the Northwest Atlantic: The 500-year perspective," J. Northwest Atl. Fish. Sci. **23**, 41–73.

Luczkovich, J. J., Mann, D. A., and Rountree, R. A. (**2008**). "Passive acoustics as a tool in fisheries science," Trans. Am. Fish. Sci. **137**, 533–541.

Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P., and Matsumoto, H. (**2007**). "An overview of fixed passive acoustic observation methods for cetaceans," Oceanography **20**, 36–45.

Midling, K. J., Soldal, A. V., Fosseidengen, J. E., and Øvredal, J. T. (**2002**). "Calls of the Atlantic cod: Does captivity restrict their vocal repertoire?," Bioacoustics **12**, 233–235.

Moore, S. E., Stafford, K. M., Mellinger, D. K., and Hildebrand, J. A. (**2006**). "Listening for large whales in the offshore waters of Alaska," BioScience **56**, 49–55.

Morano, J. L., Rice, A. N., Tielens, J. T., Estabrook, B. J., Murray, A., Roberts, B. L., and Clark, C. W. (**2012**). "Acoustically detected year-round presence of right whales in an urbanized migration corridor," Conserv. Biol. **26**, 698–707.

Nordeide, J. T., and Kjellsby, E. (**1999**). "Sound from spawning cod at their spawning grounds," ICES J. Mar. Sci. **56**, 326–332.

Rountree, R. A., Gilmore, R. G., Goudey, C. A., Hawkins, A. D., Luczkovich, J. J., and Mann, D. A. (**2006**). "Listening to fish: Applications of passive acoustics to fisheries science," Fisheries **31**, 433–446.

Rowe, S., and Hutchings, J. A. (**2006**). "Sound production by Atlantic cod during spawning," Trans. Am. Fish. Sci. **135**, 529–538.

Rowell, T. J., Schärer, M. T., Appeldoorn, R. S., Nemeth, M. I., Mann, D. A., and Rivera, J. A. (**2012**). "Sound production as an indicator of red hind density at a spawning aggregation," Mar. Ecol.: Prog. Ser. **462**, 241–250.

Schärer, M. T., Rowell, T. J., Nemeth, M. I., and Appeldoorn, R. S. (**2012**). "Sound production associated with reproductive behavior of Nassau grouper *Epinephelus striatus* at spawning aggregations," Endangered Species Res. **19**, 29–38.

Serchuk, F. M., and Wigley, S. E. (**1992**). "Assessment and management of the Georges Bank cod fishery: An historical review and evaluation," J. Northwest Atl. Fish. Sci. **13**, 25–52.

Slabbekoorn, H., Bouton, N., van Opzeeland, I., Coers, A., ten Cate, C., and Popper, A. N. (**2010**). "A noisy spring: The impact of globally rising underwater sound levels on fish," Trends Ecol. Evol. **25**, 419–427.

Urazghildiiev, I. R., and Clark, C. W. (**2007a**). "Acoustic detection of North Atlantic right whale contact calls using spectrogram-based statistics," J. Acoust. Soc. Am. **122**, 769–776.

Urazghildiiev, I. R., and Clark, C. W. (**2007b**). "Detection performances of experienced human operators compared to a likelihood ratio based detector," J. Acoust. Soc. Am. **122**, 200–204.

Urazghildiiev, I. R., Clark, C. W., Krein, T. P., and Parks, S. E. (**2009**). "Detection and recognition of North Atlantic right whale contact calls in the presence of ambient noise," IEEE J. Ocean. Eng. **34**, 358–368.

Vester, H. I., Folkow, L. P., and Blix, A. S. (**2004**). "Click sounds produced by cod (*Gadus morhua*)," J. Acoust. Soc. Am. **115**, 914–919.

Wall, C. C. (**2014**). "The benefits and challenges of passive acoustic monitoring of fish," J. Acoust. Soc. Am. **135**, 2152.

Wall, C. C., Simard, P., Lembke, C., and Mann, D. A. (**2013**). "Large-scale passive acoustic monitoring of fish sound production on the West Florida Shelf," Mar. Ecol.: Prog. Ser. **484**, 173–188.

Zelick, R., Mann, D. A., and Popper, A. N. (**1999**). *Acoustic Communication in Fishes and Frogs. Comparative Hearing: Fish and Amphibians* (Springer, New York), pp. 363–411.

Zemeckis, D. R., Hoffman, W. S., Dean, M. J., Armstrong, M. P., and Cadrin, S. X. (**2014**). "Spawning site fidelity by Atlantic cod (*Gadus morhua*) in the Gulf of Maine: Implications for population structure and rebuilding," ICES J. Mar. Sci. **71**, 1356–1365.